



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 110392

TO: Rita Mitra
Location: cm1/9b03
Art Unit: 1653
Monday, December 15, 2003

Cas Serial Number: 09/874062

From: Barb O'Bryen
Location: Biotech-Chem Library
CM1-6A05
Phone: 308-4291 *BOB*

barbara.obryen@uspto.gov

Search Notes

O'Bryen, Barbara

Fr m: Mitra, Rita
Sent: Friday, December 12, 2003 5:05 PM
T : Chan, Christina
Cc: O'Bryen, Barbara
Subj ct: Frame search request 09/874062

Hi Christina Chan

I need a frame search done foe count Mon. I have spoken to Barbara at STIC, she said it is possible for her to do. Could you please approve this rush request?

Thanks.

Rita

Please send the approval directly to Barbara O'Bryen.

I would like to request an expedited sequence search for application 09/874062 because this is an amended case and due this Monday. Please search commercial databases and published patent databases.

Please do only a **frame search** of amino acid sequence of SEQ ID NO: 3 against nucleic acid databases.

09/874062

SEQ ID NO: 3 against nucleic acid databases (**frame Search**)

Please print out results.

Thank You.

Rita

Rita Mitra, Ph. D.
Examiner
Art Unit 1653
CM-1, 9B03
(703)605-1211

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 13, 2003, 05:42:15 ; Search time 2158 Seconds
(without alignments)
2590.375 Million cell updates/sec

Title: US-09-874-062-3
Perfect score: 1484
Sequence: 1 MVSSCGSVCSQSGQGLG.....PTTCRTTCFHPICGSSCC 230

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09874062/runat_12122003_171323_24942/app_query.fasta_1.391
-DB=EST -QFMT=fastcap -SUFFIX=first -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptlo -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09874062@cgn_1_2810_ornat_12122003_171323_24942 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_eathum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	996.5	67.1	920	11 AK004258	Mus muscu
2	982.5	66.2	927	14 BY704501	BY704501
3	881	59.4	725	13 BQ193889	BY704501
4	833.5	56.2	799	10 BE250684	BE250684
5	818.5	55.2	647	28 AZ961199	2M0229J04
6	810.5	54.6	821	11 AK009035	Mus muscu
7	805	54.2	1146	11 AK029194	Mus muscu
8	790	53.2	495	13 BX283243	BX283243
9	766.5	51.7	978	11 AK009665	Mus muscu
10	762.5	51.4	705	14 BY723070	BY723070
11	744	50.1	989	14 BY704243	BY704243
12	744	50.1	1044	11 AK003994	Mus muscu
13	740	49.9	960	14 BY704302	BY704302
14	738	49.7	963	11 AK004055	Mus muscu
15	709	47.8	970	14 BY717250	BY717250
16	709	47.8	1097	11 AK017437	Mus muscu
17	694	46.8	608	10 BB613721	BB613721
18	676.5	45.6	663	14 CD052043	PLY278 Ca
19	664	44.7	692	12 BG873236	BG873236
20	656	44.2	690	28 BH050791	BH050791
21	652.5	44.0	984	11 AK020700	Mus muscu
22	637	42.9	421	9 AA500406	vi81a10.r
23	633.5	42.7	1307	11 AK079449	Mus muscu
24	632.5	42.6	428	9 AI892655	ms17f11.y
25	630.5	42.5	751	12 BG852339	602784018
26	629.5	42.4	446	9 AA760391	AA760391
27	626	42.2	1089	11 AK020690	Mus muscu
28	617.5	41.6	666	10 BB630069	BB630069
29	607.5	40.9	668	14 CD051880	PLY115 Ca
30	607	40.9	448	9 AA647179	AA647179
31	596	40.2	666	28 AZ952716	AZ952716
32	594.5	40.1	524	9 AA530660	vj44b12.r
33	594.5	40.1	713	28 AZ346186	AZ346186
34	582	39.2	379	9 AA760314	AA760314
35	561	37.8	401	9 AA760130	AA760130
36	560	37.7	437	9 AA929999	AA929999
37	554.5	37.4	388	9 AA611685	AA611685
38	554	37.3	383	9 AA856003	AA856003
39	553.5	37.3	388	9 AA154622	AA154622
40	546.5	36.8	995	9 AV089658	AV089658
41	543.5	36.6	409	9 AA791864	AA791864
42	542.5	36.6	632	14 BY723058	BY723058
43	541	36.5	702	28 AQ781300	AQ781300
44	537.5	36.2	430	9 AA672796	AA672796
45	534	36.0	466	9 AA500443	AA500443

ALIGNMENTS

RESULT 1
AK004258
LOCUS
DEFINITION
AK004258
Mus musculus 18-day embryo whole body cDNA, RIKEN full-length
enriched library, clone:110054P19 product:similar to KERATIN
ASSOCIATED PROTEIN 4.15 (FRAGMENT) [Homo sapiens], full insert
sequence.
ACCESSION
AK004258
VERSION
AK004258.1 GI:12835375
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
920 bp mRNA linear HTC 05-DEC-2002


```

Db      255  ATTTCCAGCTGCTGAGGCCACCTGCTGCGGCCACCTGCTGTATTTCCAGCTGCTGC 314
Qy      101  ArgThrThrCysCysArgProSerCysCysAlleserSerCysCysArgProSerCysCys 120
Db      315  AGGCCACCTGCTGCTGCGGCCAGCTGCTGCAATTTCCAGCTGCTGCAGC----- 362
Qy      121  IleSerSerCysCysLysProSerCysCysArgThrThrCysCysArgProSerCysCys 140
Db      363  ---GCCACCTGCTGCGGCCAGCTGCTGTAATTTCCAGCTGCTGCAGCCT----- 410
Qy      141  IleSerSerCysCysArgProSerCysCysAlleserSerCysCysLysProSerCysCys 160
Db      411  -----TCTTGCTGCGGCCAGCTGCTGTAATTTCCAGCTGCTGCAGCCT----- 464
Qy      161  GlnThrThrCysCysArgProSerCysCysAlleserSerCysCysLysProSerCysCys 180
Db      465  CGCCCGAGCTGCTGCGAGACCTAGCTGTTGCAATTTCTAGCTGCTGCCGCCAGCTGCTGT 524
Qy      181  GlnProSerCysCysArgProAlaCysCysAlleserSerCysCysHisProSerCysCys 200
Db      525  GTGTCCAGCTGCTGCAGAACCCAGTGTGCAATCTCCAGCTGCTGCCGCCCATCTGT--- 581
Qy      201  ValSerSerCysCysArgProPheSerCysProThrThrCysCysArgThrThrCysPhe 220
Db      582  -----TGCCAGACCACTGCTGCAGGACCACTGCTGCTAC 614
Qy      221  HisProIleCysCysGlySerSerCysCys 230
Db      615  CGCCCGAGCTGCTGCTAGTGTCTTGTGCTGC 644

RESULT 3
BQ193889
LOCUS      BQ193889              725 bp      mRNA      linear      EST 30-APR-2002
DEFINITION      UI-R-CNI-clz-a-10-0-UI.s1 UI-R-CNI Rattus norvegicus cDNA clone
ACCESSION      BQ193889
VERSION        BQ193889.1 GI:20369440
KEYWORDS
SOURCE        Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE      1 (bases 1 to 725)
AUTHORS        Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLES         Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL        Genome Res. 6 (9), 791-806 (1996)
MEDLINE        97044477
PUBMED        889548
COMMENT        Contact: Soares, MB
                Coordinated Laboratory for Computational Genomics
                University of Iowa
                375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
                Tel: 319 335 8250
                Fax: 319 335 9565
                Email: bent-soares@uiowa.edu
                The sequence contained an oligo-dT track that was present in the
                oligonucleotide that was used to prime the synthesis of first
                strand cDNA and therefore this may represent a bonafide poly A
                tail. The sequence tag present in the cDNA between the NotI site
                and the oligo-dT track served to identify it as a clone from the
                normalized brown adipose library cDNA Library Preparation: M.B.
                Soares lab Clone distribution: clones will be available through
                Research Genetics (www.resgen.com) The following repetitive
                elements were found in this cDNA sequence: 1-33
                >POLY.#Simple_repeat 391-516, >(CAG)n#Simple_repeat
                Seq primer: M13 Forward
                POLYA=Yes.
                Location/Qualifiers
                1..725
                /organism="Rattus norvegicus"
                /mol_type="mRNA"

```

```

/strain="Sprague-Dawley"
/db xref="taxon:10116"
/clone="UI-R-CNI-clz-a-10-0-UI"
/dev stage="adult"
/lab host="DH10B (Life Technologies)"
/clone_lib="UI-R-CNI"
/notes"Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CNI
library is a subtracted library derived from the following
pool of seven normalized rat libraries: normalized rat
seminal vesicles, normalized rat penis, normalized rat
bladder, normalized rat cervix, normalized rat brown
adipose, normalized rat fundus, and normalized rat
salivary gland. It was constructed according to the
procedure described by Bonaldo, Lennon & Soares (Genome
Research Genome 6: 791-806, 1996). For construction of
the CNI library, plasmid DNA from the pool of normalized
libraries was electroporated into competent bacteria for
the production of single-stranded circular DNA. This was
then used as a tracer in a subtractive hybridization with
a driver (PCR amplified inserts from a plamid DNA template
preparation) comprising: a) a pool of about 34,000 clones
from the Rat Unigene Set corresponding to plates R-5-AA-NN
excluding plates R-5-MM and MN. This pool represented 40%
of the final driver population. b) a pool of about 29,000
clones from subtracted libraries CA0 and CA1 corresponding
to plates R-CA0-AWV through R-CA0-AXS, R-CA0-AZX through
R-CA0-BAZ, R-CA0-BFE through R-CA0-BHY, R-CA0-BJS,
R-CA0-BKE, R-CA0-BKG-H, R-CA0-BKJ-K, R-CA0-BKP through
R-CA0-BKS, R-CA0-BKU-V, R-CA0-BLY through R-CA0-BNA,
R-CA0-BMC through R-CA0-BME, R-CA0-BNS, R-CA0-BOB through
R-CA0-BOJ, R-CA0-BPA through R-CA0-BEG, R-CA0-BBA through
R-CA0-BDA, R-CA0-BHZ through R-CA0-BJF, R-CA0-BJR,
R-CA0-BJT through R-CA0-BKE, R-CA0-BKF, R-CA0-BKI,
R-CA0-BKT, R-CA0-BKT, R-CA0-BKF, R-CA0-BKH through
R-CA0-BLN, R-CA0-BLS, R-CA0-BLF, R-CA0-BLH through
R-CA0-BLE. The resulting pool represented 20% of the
final driver population. c) a pool of about 15,000 clones
from non-normalized libraries CS0s, CT0s, CU0s, CW0s, CX0s
and normalized libraries CS0, CT0, CU0, CW0, and CX0
corresponding to plates R-CS0s-CBD through R-CS0s-CBO,
R-CT0s-CAM through R-CT0s-CAX, R-CU0s-CBP through
R-CU0s-CCA, R-CW0s-CCB through R-CW0s-CCM, R-CX0s-CCN
through R-CX0s-CCX, R-CS0-BSD, R-CS0-BTD through R-CS0-BTV
, R-CS0-BVM, R-CT0-BTW through R-CT0-BUP, R-CT0-BVN,
R-CU0-BUQ through R-CU0-BVL, R-CW0-BVV through R-CW0-BWP,
R-CW0-BXN through R-CW0-BXO, R-CX0-BMQ through R-CX0-BXM.
The resulting pool represented 5% of the final driver
population. d) a pool of about 5,000 clones (1,000 from
non-normalized eye library CV0 and 4,000 from normalized
eye library CV1) corresponding to plates R-CV0-BRH through
R-CV0-BRR, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through
R-CV1-BTC, and R-CV1-BVO through R-CV1-BVU. This pool
represented about 5% of the final driver population. e) A
pool of about 10,000 clones from subtracted library BS2,
BV0 and BV0p (7-9.5 kb cDNA library fraction from rat
whole embryo), and BX0 (0.5-7kb cDNA library fraction from
rat whole embryo) corresponding to plates R-BS2-BDB
through R-BS2-BFB, R-BV0-ANK through R-BV0-ANR, R-BV0p-AOI
through R-BV0p-AOX, and R-BX0-AQY through R-BX0-ASH. The
resulting pool represented 5% of the final driver
population. f) a pool of about 7,000 clones from the
seven non-normalized libraries that make up the tracer
including CY0, CZ0, DA0, DB0, DC0, DD0, and DE0
corresponding to plates R-CY0-BXP through R-CY0-BXZ,
R-CZ0-BYA through R-CZ0-BYI, R-CZ0-BZB-C, R-DA0-BYJ
through R-DA0-BYP, R-DA0-BZD through R-DA0-BZH, R-DB0-BYQ
through R-DB0-BZA, R-DC0-BZI through R-DC0-BZQ, R-DD0-CAY
through R-DD0-CBA, R-DD0-BZK through R-DD0-CAA, The
R-DD0-CBB-C, and R-DE0-CAB through R-DE0-CAL. The
resulting pool represented about 10% of the final driver
population. g) a pool of about 2,000 clones from the pool
of normalized libraries, CN0, that makes up the tracer.

```


Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5', GAGAGAGAAGATCAACGAGCTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using tris(halo) thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer added of sequence [5', GAGAGAGAGATCTTCGATCTTAATTAATATCCCTCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

FEATURES

Location/Qualifiers

```
1. .978
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:2310037K05"
/db_xref="MGI:1909881"
/db_xref="taxon:10090"
/clone="2310037K05"
/sex="male"
/tissue_type="tongue"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
55. 561
```

CDS

```

/notes="unnamed protein product; putative
similar to KERATIN ASSOCIATED PROTEIN 4.15 (FRAGMENT)
[ Homo sapiens] (SPRR|Q9BY05, evidence: FASTY, 70.7%ID,
96.8%length, match=495)"
/codon_start=1
/protein_id="BAB26426.1"
/db_xref="GI:112844600"
/db_xref="MGI:19234594"
/translations="MWSSSCGSVSCSEGGSCQCPSCCVSSCRPCCQSVCCQPTQ
CRPSCTISCCRSPCCRSCCRPCCQSCVSSCCRPCCQACQPTCCRSPCCRSPCCIS
SCQPCSGGSSCGGSSCCRPCCRPCCCLRPVCGVCCQCTTCVPTCVISTCPREPR
MCCATPCC"
a      328 c      212 g      270 t

4.14e-45      Length:      978
766.50      Matches:      120
72.63%      Conservative:      18
63.16%      Mismatches:      34
51.65%      Indels:      18
11      Gaps:      4

```

US-09-874-062-3 (1-230) x AK009665 (1-978)

Qy	1	MetValSerSerCysCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGly	20
Db	55	ATGCTCAGCTCCTGTGTGGCTCTGCTCTCAGGAGGGCTGTAGCCAAGGC	108
Qy	21	GlnGluSerCysCysArgProSerCysCysGlnThrThrCysCysArgThrThrCysCys	40
Db	109	-----TGCTGCCAGCCAGCTGCTGTGTCTCAGCTGCTGCAGGCGCTCAGTGCTGC	159
Qy	41	ArgProSerCysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCys	60
Db	160	CAGTCTGTGCTGCTGCCAGCCACCTGCTGCCGCCCGCCAGCTGCTGCATCTCAGCTGCTGT	219
Qy	61	LysProSerCysCysLeuThrThrCysCysArgThrThrCysCysArgProSerCysCys	80
Db	220	CGCCCCAGC-----TGCTGTAGGCCCATGTTGCTGCAGGCCCCAGCTGCTGT	264
Qy	81	IleSerSerCysCysArgProSerCysCysIleSerSerCysCysLysProSerCysCys	100
Db	265	GTGTCCAGCTGCTGCAGACCCAGTGCTGCCAGTCTGCGTGTGCCAGCCCACTGCTGC	324
Qy	101	ArgThrThrCysCysArgProSerCysCysIleSerSerCysCysArgProSerCysCys	120
Db	325	CGCCCCAGCTGCTGCCGCCAGCTGCTGCATCTCAGCTGCTGCCAACCATCTGTGTGT	384

Qy	121	IleSerSerCysCysIysIysProSerCysCysArgThrThrCysCysArgProSerCysCys	140
Db	385	GGTCTAGCTGCTGTGCTCCAGCTCTGCGCGC---CCTTGTGCGCGCGCCC-----	432
Qy	141	IleSerSerCysCysArgProSerCysCysIleSerSerCysCysIysProSerCysCys	160
Db	433	-----TGTTCGCGCCCTGCTGCTGTCTCTGAGACCACTCTGCGGTCAAGGTCTGCTGC	483
Qy	161	GlnThrThrCysCysArgProSerCysCysIleSerSerCysThrArgProGlnCysCys	180
Db	484	CAAAACCACTTGCTACCGCCCACTTGTGTGCATCTCCACCTGCCTCGCCCATGTGCTGT	543
Qy	181	GlnProSerCysCys-ArgProAlaCys	189
Db	544	GCCACCCCTGTTGCTGAGCCTCCATGC	571
RESULT 10			
BY723070			
LOCUS			
DEFINITION			
	BY723070	705 bp	linear
	BY723070	RIKEN full-length enriched, 6 days neonate	EST 17-DEC-2002
	musculus cDNA clone A03010k20 5',	RNA sequence.	skin Mus

ACCESSION
VERSION
KEYWORDS

SOURCE ORGANISM

.....

REFERENCE AUTHORS

CONCLUSIONS

TITLE

JOURNAL

COCHRANE
MEDLINE
PUBMED

COMMENT

QY 175 TyrArgProGlnCysGlnProSerCysCysArgProAlaCysCysIleSerSerCys 194
 Db 506 CTCAC---CTGTGTCGCAGCCAGCTGCTGTGGACCTGTGACCTGCACCAAGGACTTGC 562
 QY 195 CysHisProSerCys---CysValSerSerCys-----ArgCysProPheSerCys 210
 Db 563 TACCAGCCCAACATGTGTCTGTGCTGGCTGCCTGTCCCAAGGCTGTGGGTCTTAACCTGC 622
 QY 211 ProThrThrCysCys 215
 Db 623 TCCCAATCTGTGTGT 637
 RESULT 13
 BY704302
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
 Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamaguchi,I., Kiyosawa,H.,
 Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
 Gojohori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
 Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
 Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,
 L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,
 A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.,
 Gough,J., Grimmond,S., Guetincich,S., Hirokawa,N., Jackson,I.J.,
 Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
 King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,
 P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,
 H., Nagashima,T., Numa,K., Okido,T., Pavan,W.J., Perle,G.,
 Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
 Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,
 B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,
 M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,
 R.D., Tomita,M., Verdaro,R., Wagner,L., Wahlestedt,C., Wang,Y.,
 Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,
 M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
 Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,
 M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
 Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,
 Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,
 K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,
 E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 22354683
 PUBLISHED
 COMMENT
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gscc.riken.go.jp/
 URL:http://genome.gsc.riken.go.jp/
 Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Hirozane,T., Hori,F., Imotani,K.,
 S., Hashizume,W., Hayashida,K., Kawai,J., Kojima,Y., Kondo,S., Konno,
 Ishii,Y., Itoh,M., Kagawa,I., Murata,M., Nakamura,M., Nomura,K.,
 H., Koya,S., Miyazaki,A., Murata,M., Saito,R., Sakazume,N., Sano,H.,
 Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
 Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
 Waki,K., Watanishi,A., Muramatsu,M. and Hayashizaki,Y. Direct
 Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.
 Location/Qualifiers
 1. 960
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="111003F04"
 /tissue_type="whole body"
 /dev_stage="18-day embryo"
 /clone_lib="RIKEN full-length enriched, 18-day embryo
 whole body"
 BASE COUNT 185 a 308 c 196 g 271 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,086-43 Length: 960
 Score: 740.00 Matches: 119
 Percent Similarity: 73.26% Conservative: 18
 Best Local Similarity: 63.64% Mismatches: 28
 Query Match: 49.87% Indels: 22
 DB: 14 Gaps: 8
 US-09-874-062-3 (1-230) x BY704302 (1-960)
 QY 1 MetValSerSerCysCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGly 20
 Db 62 ATGGTCACCTCT---TGTGGCTCTGTCTGCTCTGAAAGGGCTGTGGCCAGGC----- 112
 QY 21 GlnGluSerCysCysArgProSerCysCysGlnThrThrCysCysArgThrThrCysCys 40
 Db 113 -----TGTGTCAGCCAGCTGCTGTCACCAACCACTGCTGTAGGACCACTGCTGTC 163
 QY 41 ArgProSerCysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCys 60
 Db 164 CGCCCCAGCTGCTGTGTGTCAGCTGCTGTCAGGCCAGCTGCTGTGTGTCAGCTGCTGTC 223
 QY 61 LysProSerCysCysLeuThrThrCysCysArgThrThrCysCysArgProSerCysCys 80
 Db 224 AGGCCAGCTGCTGTCAGGCCAGCTGCTGTCAGCTGCTGTCAGCTGCTGTCAGCTGCTGTC 274
 QY 81 IleSerSerCysCysArgProSerCysCysIleSerSerCysCysArgProSerCysCys 100
 Db 275 -----ACCTGTCGCCGCCAGCTGCTGTCATTTTCAGCTGCTGTCACCACTGCTGTCGT 328
 QY 101 ArgThrThrCysCysArgProSerCysCysIleSerSerCysCysArgProSerCysCys--- 119
 Db 329 AGCTCAGCTGCTGTGGTGTCTAGTTGCTGTC---CGCCCCCTGCTGCGCCCTGCTGTCAGC 385
 QY 120 ---CysIleSerSerCysCysArgProSerCysCysArgThrThrCysCysArgProSer 138
 Db 386 CCTTGCTGACGCCCTGTTGGCTGCC---TGCTGCCGA---CCCTGTCGCCGCCCC--- 436
 QY 139 CysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCysArgProSer 158
 Db 437 -----TGTTGCCGCCCTGCTGCTGCTGAGAACAGTCTGTGTTGGTTCAGGTC 481

ORIGIN

Alignment Scores:
 Pred. No.: 4,298-43 Length: 963
 Score: 738.00 Matches: 119
 Percent Similarity: 72.73% Conservative: 17
 Best Local Similarity: 63.64% Mismatches: 29
 Query Match: 49.73% Indels: 22
 DB: 11 Gaps: 8

US-09-874-062-3 (1-230) x AK004055 (1-963)

Qy 1 MetValSerSerCysCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGly 20
 Db 62 ATGGTCAGCTCT--TGTGGCTCTGCTGCTCTGAAAGGGCTGTGGCCAGGC----- 112

Qy 21 GlnGluSerCysCysArgProSerCysCysGlnThrThrCysCysArgThrThrCysCys 40
 Db 113 -----TGCTGGCAGCCCACTGCTGCCAACCCACCTGCTGTAGGACCACCTGCTGTC 163

Qy 41 ArgProSerCysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCys 60
 Db 164 CCCCCCAGCTGCTGTGTCAGCTGCTGCAGGCCAGCTGCTGTGTCCAGCTGCTGTC 223

Qy 61 LysProSerCysCysLeuThrThrCysCysArgThrThrCysCysArgProSerCysCys 80
 Db 224 AGCCCAAGTTGCTGCAGGCCCACTGCTGCCAGTCTGTGTGTCGACCGC----- 274

Qy 81 IleSerSerCysCysArgProSerCysCysIleSerSerCysCysArgProSerCysCys 100
 Db 275 -----ACCTGTGCGCGCCCACTGCTGCTCAATTCAGCTGCTGTCAACCCCTCTGTGT 328

Qy 101 ArgThrThrCysCysArgProSerCysCysIleSerSerCysCysArgProSerCysCys 119
 Db 329 AGCTCCAGCTGCTGTGTGTGTAGTTGCTGC-----CCTGCTGCTGCGCGCCCGC 385

Qy 120 ---CysIleSerSerCysCysLysProSerCysCysArgThrThrCysCysArgProSer 138
 Db 386 CTTGTGCTGACGCGCCCTGTGTGCGTCCC---TGTGCGCGA---CCCTGCTGCGCGCCCGC 436

Qy 139 CysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCysLysProSer 158
 Db 437 -----TGTTGCGCGCCCTGCTGCTGTGCTGAGACCACTGCTGTGCTGAGTGC 481

Qy 159 CysCysGlnThrThrCysCysArgProSerCysCysIleSerSerCysCysArgProGln 178
 Db 482 TGCTGTCACCAACCACTTGTACCGCCCACTGCTGTGTCATCTCCACCTGCGCCCGCCCATG 541

Qy 179 CysCysGlnProSerCysCys 185
 Db 542 TGCTGTGCCATCCCTGCTGTC 562

RESULT 15
 BY171250
 LOCUS
 DEFINITION
 BY171250 RIKEN full-length enriched, 10 days neonate head Mus
 musculus cDNA clone 530401L02 5', mRNA sequence.

ACCESSION
 VERSION
 BY171250
 KEYWORDS
 EST.

SOURCE
 Mus musculus (house mouse)

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 970)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H.,
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,
 Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A.,
 Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S.,
 Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani,
 L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. P., Forrest,
 A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.

Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J.,
 Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M.,
 King, B. L., Konegaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons,
 P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki,
 H., Nagashima, T., Numata, K., Okido, T., Pavan, M. J., Pertea, G.,
 Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ring,
 Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring,
 B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou,
 M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale,
 R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y.,
 Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa,
 M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,
 Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura,
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Kondo, S.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, M., Imotani, K., Ishii,
 Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,
 K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,
 E. S., Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 22354683
 12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp

URL: <http://genome.gsc.riken.go.jp/>

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda,
 S., Hashizume, M., Hayashida, K., Hirozane, T., Hori, F., Imotani, K.,
 Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno,
 H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K.,
 Numazaki, R., Ohno, M., Osato, N., Saito, R., Sakazume, N., Sano, H.,
 Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y.,
 Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct
 Submission

Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.

FEATURES
 source

Location/Qualifiers
 1..970
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="530401L02"
 /sex="mixed"
 /tissue_type="head"
 /dev_stage="10 days neonate"
 /lab_host="DH10B"
 /clone_lib="RIKEN full-length enriched, 10 days neonate
 head"

Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATTCGAGATTAAATTAATCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
BamHI."

BASE COUNT 210 a 327 c 202 g 230 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 4.94e-41 Length: 970
Score: 709.00 Matches: 123
Percent Similarity: 59.09% Conservative: 20
Best Local Similarity: 50.83% Mismatches: 50
Query Match: 47.78% Indels: 50
DB: 14 Gaps: 12

US-09-874-062-3 (1-230) x BY717250 (1-970)

```
QY 1 MetValSerSerCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGly 20
Db 53 ATGACCAACTCTCTGCTGC----- 70

QY 21 GlnGluSerCysCysArgProSerCysGlnThrCysCysArgThrCysCys 40
Db 71 ---TCCCTTGTGCGACCCACCTGCTGCGAGCACCTGCTGCGAGCACCTGCTGG 127

QY 41 ArgProSerCys-----CysIleSerSerCysCysArgProSerCysIleSer 57
Db 128 AGACCAAGCTGTGTGACCAAGCTGTTGCCAGGCCCTGCTGCCAGCCAGCTGTGTGGTCC 187

QY 58 SerCysCysLysProSerCysCysLeuThrCysCysArgThr-----ThrCys 74
Db 188 AGCTGTGCGACGCT---TGCTGTCAACACCTGCTGCGAGACCTGCTTCCAGCCATGC 244

QY 75 CysArgProSerCysCysIleSerSerCysCysArgProSerCysCysIleSerSerCys 94
Db 245 TGTGTGAGCAGCTGTGCGCACACCTGCTGCCAGCCCTGCTGTGTGTCTCCAGCTGC 304

QY 95 CysLysProSerCysCysArgThrCysCysArgProSerCysCysIleSerSerCys 114
Db 305 TGCCAGCCC-----TGCTGCCAGCCAGCTGCTGTGTCAGTCCAGCTGC 346

QY 115 CysArgProSerCysCysIleSerSerCysCysLysProSerCysCysArgThrCys 134
Db 347 TGCCAGCCAGCTGTGCGAGCCAGCTGCTGCCAGCCAGCTGCTGTGTCAGCCAGCTGC 406

QY 135 CysArgProSerCysCysIleSerSerCysCysArgProSerCysCysIleSerSerCys 154
Db 407 TGTATCTCAGCTGCTGC---CAGCCCTGCTGCGAGCCCTAGCTGTGTCAGTCCAGCTGC 463

QY 155 CysLysProSerCys-----CysGlnThrCysCysArgProSerCys----- 169
Db 464 TGCAAGCCCTGCTGCCAGCCCTTCTGCTCTCAACCTGCTGTGTCAGCCAGCTGCTCTGGA 523

QY 170 -----CysIleSerSerCysTyrArgProGlnCys---CysGlnProSerCysCys 185
Db 524 CTGTGACCTGCACCGAGACTTGCTACCGCCACCTGCTGCTGTGCTGCTGCTGCTG 583

QY 186 ArgProAlaCysCysIleSerSerCysCysHisProSerCysCysValSerSerCysArg 205
Db 584 TCCCAAGGCTGTGGG---TCCAGCTGCTGTGAGCCC-----TGTCGC 622
```

```
QY 206 CysProPheSerCysProThrThr-----CysCysArgThrThrCysPheHis 221
Db 623 TGTGTG-ATCAACTTACCAGACCCACTAGCAACAATCTGCAAGATAACCAATACAT 681

QY 222 Profile 223
Db 682 CAAATC 687
```

Search completed: December 13, 2003, 07:30:04
Job time : 2165 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 13, 2003, 05:47:02 ; Search time 331 Seconds
(without alignments)
2309.450 Million cell updates/sec

Title: US-09-874-062-3

Perfect score: 1484

Sequence: 1 MVSSCCSVCSQSGGLG.....PTTCRTTCFHPICCGSSCC 230

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09874062/runat 12122003 171324 24977/app query.fasta_1.391
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=cblosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09874062 @CIGN 1 1 352 @runat 12122003 171324 24977
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2.*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description

1	1484	100.0	1101	9	US-09-874-062-2	Sequence 2, Appli
2	1053	71.0	911	13	US-10-120-988-43	Sequence 43, Appl
3	762.5	51.4	469	13	US-10-029-386-20613	Sequence 20613, A
4	733	49.4	3489	13	US-10-294-804-1	Sequence 1, Appli
5	671.5	45.2	1229	13	US-10-029-386-25152	Sequence 25152, A
6	666.5	44.9	1980	13	US-09-950-051-6	Sequence 6, Appli
7	661.5	44.6	1196	13	US-10-029-386-20464	Sequence 20464, A
8	643.5	43.4	1002	13	US-10-029-386-20498	Sequence 20498, A
9	643.5	43.4	1002	13	US-10-029-386-25183	Sequence 25183, A
10	615.5	41.5	898	13	US-10-027-632-284090	Sequence 284090, A
11	615.5	41.5	898	14	US-10-027-632-284090	Sequence 7, Appli
12	603	40.6	1800	13	US-09-950-051-7	Sequence 6314, Appli
13	602.5	40.6	478	9	US-09-864-761-6314	Sequence 24945, A
14	601	40.5	865	13	US-10-029-386-24945	Sequence 24945, A
15	601	40.5	1974	9	US-09-864-761-4747	Sequence 4747, Ap
16	589.5	39.7	915	13	US-10-029-386-20331	Sequence 20331, A
17	574	38.7	489	9	US-09-864-761-27528	Sequence 27528, A
18	570.5	38.4	461	9	US-09-864-761-10890	Sequence 10890, A
19	561.5	37.8	709	13	US-10-027-632-110242	Sequence 110242, A
20	561.5	37.8	709	13	US-10-027-632-110243	Sequence 110243, A
21	561.5	37.8	709	13	US-10-027-632-110244	Sequence 110244, A
22	561.5	37.8	709	14	US-10-027-632-110242	Sequence 110242, A
23	561.5	37.8	709	14	US-10-027-632-110243	Sequence 110243, A
24	561.5	37.8	709	14	US-10-027-632-110244	Sequence 110244, A
25	552.5	37.2	765	13	US-10-029-386-24902	Sequence 24902, A
26	550	37.1	11872	10	US-09-764-847-1390	Sequence 1390, Ap
27	550	37.1	11872	15	US-10-092-154-1390	Sequence 1390, Ap
28	544.5	36.7	615	13	US-10-029-386-25308	Sequence 25308, A
29	523.5	35.3	806	13	US-10-120-988-427	Sequence 427, App
30	513	34.6	696	9	US-09-864-761-21486	Sequence 21486, A
31	508	34.2	507	13	US-10-029-386-25198	Sequence 25198, A
32	508	34.2	538	13	US-10-027-632-273487	Sequence 273487, A
33	508	34.2	538	13	US-10-027-632-273489	Sequence 273489, A
34	508	34.2	538	13	US-10-027-632-273490	Sequence 273490, A
35	508	34.2	538	14	US-10-027-632-273487	Sequence 273487, A
36	508	34.2	538	14	US-10-027-632-273489	Sequence 273489, A
37	508	34.2	538	14	US-10-027-632-273490	Sequence 273490, A
38	507	34.2	538	13	US-10-027-632-273488	Sequence 273488, A
39	507	34.2	538	14	US-10-027-632-273488	Sequence 273488, A
40	500	33.7	1179	13	US-10-027-632-263173	Sequence 263173, A
41	500	33.7	1179	14	US-10-027-632-263173	Sequence 263173, A
42	496	33.4	565	13	US-10-029-386-25172	Sequence 25172, A
43	496	33.4	581	13	US-10-029-386-20211	Sequence 20211, A
44	495	33.4	508	13	US-10-029-386-11465	Sequence 11465, A
45	489.5	33.0	2045	13	US-10-027-632-99699	Sequence 99699, A

ALIGNMENTS

RESULT 1
US-09-874-062-2
; Sequence 2, Application US/09874062
; Patent No. US20020081607A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Four Disulfide Core Domain-Containing (FDCD) Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT033P1
; CURRENT APPLICATION NUMBER: US/09/874,062
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US00/32462
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/168,229
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; TYPE: DNA
; LENGTH: 1101
; ORGANISM: Homo sapiens
US-09-874-062-2
Alignment Scores:


```
Db 716 TCACCTGTCCCGCCCTTGTGTGCTGGCTCTCTCTTGTCTGC 757
RESULT 3
US-10-029-386-20613
; Sequence 20613, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20613
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006070.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.72
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EST_HUMAN HIT: BE250684.1, EVALUE 4.00e-67
; OTHER INFORMATION: SWISSPROT HIT: P26371, EVALUE 5.50e+00
; OTHER INFORMATION: NT HIT: G114210533, EVALUE 0.00e+00
US-10-029-386-20613
Alignment Scores:
Pred. No.: 4,21e-53 Length: 469
Score: 762.50 Matches: 115
Percent Similarity: 75.43% Conservatives: 17
Best Local Similarity: 65.71% Mismatches: 24
Query Match: 51.38% Indels: 19
DB: 13 Gaps: 3
US-09-874-062-3 (1-230) x US-10-029-386-20613 (1-469)
Qy 1 MetValSerSerCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGly 20
Db 1 ATGTCAACTCTGTGTGGCTCTGTGTGCTCTGACACAGGCTGT-----GGC 48
Qy 21 GlnGluSerCysArgProSerCysGlnThrThrCysCysArgThrThrCysCys 40
Db 49 CTAGAGAACTGTCGCGTCCAGCTACTGTCAGACACCTGCTGCAGGACCACTGCTGC 108
Qy 41 ArgProSerCysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCys 60
Db 109 CGCCCCAGCTGTGTGTGCTGAGCTGCTGACAG----- 141
Qy 61 LysProSerCysCysLeuThrThrCysCysArgThrThrCysCysArgProSerCysCys 80
Db 142 ---CCCCAGTGTGCCAGACCACTGCTGAGGACCACTGCTGCCACCCAGCTGTGT 198
Qy 81 IleSerSerCysArgProSerCysCysIleSerSerCysCysLysProSerCysCys 100
Db 199 GTGTCCAGCTGTGCAGACCCAGTGTGCTGCCAGTCTGTGTGCTGCCAGCCACCTGCTGC 258
Qy 101 ArgThrThrCysCysArgProSerCysCysIleSerSerCysCysArgProSerCysCys 120
Db 259 AGACCCCAATGCTGCCAGTACTGCTGTGTAGGACCACTGCTGCCCGCCAGCTGTGCTGC 318
Qy 121 IleSerSerCysCysLysProSerCysCysArgThrThrCysCysArgProSerCysCys 140
Db 319 AGGCCCCAGTGTGCCAGTGTGTGTGCTGCCAGCCCACTGCTGTGCTGCCAGCTACTGT 378
Qy 141 IleSerSerCysCysArgProSerCysCysIleSerSerCysCysLysProSerCysCys 160
Db 1691 GGCTCCCGCTGCTGGGGCTCCCGCTGCTGGCTCCCGCTGCTGGCTCCCGCTGCTGTGT 1632
Db 1691 GGCTCCCGCTGCTGGGGCTCCCGCTGCTGGCTCCCGCTGCTGGCTCCCGCTGCTGTGT 1632
Db 379 GTGTCCAGCTGTGTCAGACCC-----CAGTGTGCCAGACCACTGTGCTGC 423
Qy 161 GlnThrThrCysCysArgProSerCysCysIleSerSerCysCysTyr 175
Db 424 AGAACCACTGTGCTGCCGCCACCTGCTGTGTGTGTCCAGAGTGTCTAC 468
RESULT 4
US-10-294-804-1/c
; Sequence 1, Application US/10294804
; Publication No. US20030133948A1
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Cotter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; TITLE OF INVENTION: to Genomic Host DNA
; FILE REFERENCE: UM-03778
; CURRENT APPLICATION NUMBER: US/10/294,804
; CURRENT FILING DATE: 2002-11-14
; PRIOR APPLICATION NUMBER: US/09/410,399
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-294-804-1
Alignment Scores:
Pred. No.: 8,92e-50 Length: 3489
Score: 733.00 Matches: 122
Percent Similarity: 57.32% Conservatives: 15
Best Local Similarity: 51.05% Mismatches: 88
Query Match: 49.39% Indels: 14
DB: 13 Gaps: 4
US-09-874-062-3 (1-230) x US-10-294-804-1 (1-3489)
Qy 3 SerSerCysCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGlyGlnGlu 22
Db 2162 TCCTGTGCTGCTCATCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2112
Qy 23 SerCysCysArg---ProSerCysCysGlnThrThrCysCysArgThrThrCysCysArg 41
Db 2111 TCCTGTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2052
Qy 42 ProSerCysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCys 61
Db 2051 TCATCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1992
Qy 62 ProSerCysCysLeuThrThrCysCysArgThrThrCysCysArgProSerCysCys 81
Db 1991 TCATCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1932
Qy 82 SerSerCysCysArgProSerCysCysIleSerSerCysCysLysProSerCysCysArg 101
Db 1931 TCATCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1872
Qy 102 ThrThrCysCysArgProSerCysCysIleSerSerCysCysArgProSerCysCys 121
Db 1871 TCATCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1812
Qy 122 SerSerCysCysLysProSer---CysCysArgThrThrCysCysArgProSerCysCys 140
Db 1811 TCATCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1752
Qy 141 IleSerSerCysCysArgProSerCysCysIleSerSerCysCysLysProSerCysCys 160
Db 1751 GGCTCCCGCTGCTGGCTCCCGCTGCTGGCTCCCGCTGCTGGCTCCCGCTGCTGTGT 1692
Qy 161 GlnThrThrCys-----CysArgProSerCysCys 171
Db 1691 GGCTCCCGCTGCTGGGGCTCCCGCTGCTGGCTCCCGCTGCTGGCTCCCGCTGCTGTGT 1632
```



```
Qy 23 SerCysCysArgProSerCysCysGlnThrThrCysCysArgThrThrCysCysArgPro 42
Db 967 AGCTCTGTCAGCGCCTCGTGTGCGCAGAGTCTAGTGTCCAGCTGGCTTGTGTGCTCC 1026
Qy 43 SerCysCysLeuSerSerCysCysArgProSerCysCysCysLeuSerSerCysCysArgPro 62
Db 1027 TCCCTTGCAGCGCCTGCTGCTGCGTGTGCTGCGAAGACTGTCTGTGCAAGCCT 1086
Qy 63 SerCysCysLeu-----ThrThrCysCysArgThrThrCysCysArg 76
Db 1087 GTGTGCTGTGCGCGCTGCTGTGCGGATTTCTTTCATGCTGCCAGCAGTCTAGTGTCCAG 1146
Qy 77 ProSerCysCysLeuSerSerCysCysArgProSerCysCysCysLeuSerSerCysCysArg 96
Db 1147 TCAGCTTGTGTCACCTCTCCCTGCGCAGCAGCGCTGTGTGTGCCCATCTGTGTCAAG 1206
Qy 97 Pro----- 97
Db 1207 CCTGTCTGTGCTGGGATTTCTTTCGTGTGCGCAGCAGTCTAGTGTGTGAGCTGTGTG 1266
Qy 98 -----SerCysCysArgThrThrCys----- 104
Db 1267 TCCAGCCCTGCTGCGCAGCGGTCTGTGAGCCAGCCCTGCCAATCAGCTGCATCAGC 1326
Qy 105 ---CysArgProSerCysCysLeuSerSerCysCysArgProSerCysCysCysLeuSerSer 123
Db 1327 TCCTGTCAGCGCCTCGTGTGCGCAGCAGTCTAGTGTGCCAGCGGCTTGTGTGCACTCTCC 1386
Qy 124 -----CysCysLeuProSerCysCysArgThrThrCysCysArgPro--- 137
Db 1387 TCCTGCCAGCAGCGCTGCTGCGTCCCGCTGTGTGCAAGACTGTCTGTGTCAAGCCTGTG 1446
Qy 138 -----SerCysCysLeuSerSerCysCysArgProSerCysCysCysLeu 151
Db 1447 TGCTCTGAGGATTTCTTTCATGCTGCCAGCAGTCTAGTGTGCCAGCGGCTTGTGTGACC 1506
Qy 152 SerSerCysCysLeuProSerCysCysGlnThrThrCysCysArgProSerCysCys--- 170
Db 1507 TCCTCTCCCTGCCAGCAGCGTGTGTGTGCTGTGTGCTGTGCAAGCCTGTGTGTGCAAG 1566
Qy 171 -----IleSerSerCysThrArgProGlnCysCysGlnProSerCys 184
Db 1567 CCTGTGCGCTGTGCGCCATCTGTCTGTGGGCTTCTCTCTGTGCTGCCAGCAGTCTAGC 1626
Qy 185 CysArgProAlaCysCysLeuSerSerCysCysHisProSerCysCysValSerSerCys 204
Db 1627 TGCAGCAGCTTGTGTCACCTCTCTCCAAAGCAGCAGGCTGTGCGTCCCGCTGTG 1686
Qy 205 ArgCysProPheSerCys-----ProThrThrCysCysArgThr 217
Db 1687 TGCAAGCTGTGAGCTGTGTGCGCTGTTTGTCTGCGGCTTCTCTTTCATGTGTGCCAGAA 1746
Qy 218 ThrCysPheHisProIleCysCysGlySerSerCysCys 230
Db 1747 TCTAGCTGCCAGCAGCTGTGTGCAACCACCTCTCTGTGC 1785
```

RESULT 7

```
US-10-029-386-20464
; Sequence 20464, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AECOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20464
; LENGTH: 1196
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AJ011930.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: NT HIT: AL163300.2, EVALUATE 1.00e-123
; OTHER INFORMATION: SWISSPROT HIT: P08131, EVALUATE 2.00e-04
; OTHER INFORMATION: EST_HUMAN HIT: BF057518.1, EVALUATE 1.00e-120
US-10-029-386-20464
```

Alignment Scores:

Pred. No.:	1-52e-44	Length:	1196
Score:	661.50	Matches:	118
Percent Similarity:	51.90%	Conservative:	32
Best Local Similarity:	40.83%	Mismatches:	76
Query Match:	44.58%	Indels:	63
DB:	13	Gaps:	12

US-09-874-062-3 (1-230) x US-10-029-386-20464 (1-1196)

```
Qy 3 SerSerCysCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGlyGlnGlu 22
Db 262 AGCTCTGTCAGCGCCTCGTGTGCCAGCAGTCTAGTGTGCCAGCTGGCTTGTGTGCTCC 321
Qy 23 SerCysCysArgProSerCysCysGlnThrThrCysCysArgThrThrCysCysArgPro 42
Db 322 TCCCTTGCAGCAGCGCTGTGTGTGCAAGACTGTCTGTGTCAAGCCT 381
Qy 43 -----SerCysCysLeuSerSerCysCysArg 51
Db 382 GTGTACTGTGTGCTGTGTGCTGAGTGGGATTTCTTTCATGCTGCCAGCAGTCTAGTGTCCAG 441
Qy 52 ProSerCysCysLeuSerSerCysCysLysProSerCysCysLeuThrThrCysCysArg 71
Db 442 TCAGCTTGTGTGACCTCTCTCCCTGCGCAGCAGCGCTGTGTGTGCCCCATCTGTGTCAAG 501
Qy 72 -----ThrThrCysCysArgProSerCysCysLeuSer--- 82
Db 502 CCTGTCTGTCTGGGATTTCTTCTGTGTGTGCCAGCAGTCTAGTGTGTGAGCTGTGTG 561
Qy 83 -----SerCysCysArgProSerCysCysCysLeuSerSerCys--- 94
Db 562 TCCAGCCCTGTGTGCCAGCGGCTGTGTGAGCCAGCCCTGCCAATCAGCTGTGATCAGC 621
Qy 95 ---CysLysProSerCysCysArgThrThrCysCysArgProSerCysCysLeuSerSer 113
Db 622 TCCTGTCAGCGCCTCGTGTGCCAGCAGTCTAGTGTGCCAGCGGCTTGTGTGACCTCTCC 681
Qy 114 CysCysArgProSerCysCysLeuSerSerCysCysLysProSerCysCysArgThrThr 133
Db 682 CCCTGTCAGCAGCGCTGTGTGTGCTGTGTGTGCAAGCCGCTGTGTGTGTGCCACC 741
Qy 134 -----CysCysArgProSerCysCysCysLeuSerSer--- 143
Db 742 TGCTGTGATGATTCGGGTTTCATGTGCCAGCAGCTTGTGTGACCTCTCTCCAAAGCCAG 801
Qy 144 -----CysCysArgProSerCysCysLeuSerSerCysCysLysPro--- 157
Db 802 CAGGGCTGTGTGCGTGTGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGG 861
Qy 158 -----SerCysCysGlnThrThrCysCysArgProSerCysCysLeuSerSerCys 174
Db 862 GCTTCCAGTTTCATGTGTGCCAGCAGTCTAGTGTGCCAGCGGCTTGTGTGACCACTCTGC 921
Qy 175 TyrArgProGln-----CysCysGlnProSerCysCysArgProAlaCys 189
Db 922 TGCAGACCCCTCTCTCTCCGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGC 978
```



```
US-09-874-062-3 (1-230) x US-10-029-386-25183 (1-1002)
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 284090
; LENGTH: 898
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-284090

Alignment Scores:
Pred. No.: 5,48e-41 Length: 898
Score: 615.50 Matches: 107
Percent Similarity: 48.11% Conservative: 33
Best Local Similarity: 36.77% Mismatches: 81
Query Match: 41.48% Indels: 71
DB: 13 Gaps: 10

US-09-874-062-3 (1-230) x US-10-027-632-284090 (1-898)
Qy 4 SerCysCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeu----- 19
Db 31 GCCTGCTGATGCGCCGCTGCTGCTGCAAGACTGTCTGCTGCAAGCGTGTGTACTGTGTGCCT 90
Qy 20 -----GlyGlnGlnSerCysCysArgProSerCysCysGlnThrCysCysArg 36
Db 91 GTCTGCAGTGGGGATTTCTTCATGCTGCCAGCAGTCTAGTGTGCCAGTGTGTGTGTGCACC 150
Qy 37 ThrThrCysCysArgProSerCysCysGlnSerCysCysArgProSerCys----- 54
Db 151 TCCTCCCTCCCTGCCAGCAGCGCTGCTGTGTGCCCATCTGTGTGCAAGCGTGTGTGTGTGG 210
Qy 55 CysIleSerSerCysCysCysArgProSerCysCysLeu-----ThrThrCysCys 70
Db 211 ATTTCTCTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 270
Qy 71 ArgThrThrCys-----CysArgProSer 78
Db 271 CAGGCGGTCTGTGAGCCAGCCCTGCTGCAATCAGGCTGTGATCAGTCTGTGCAAGCGCTGTG 330
Qy 79 CysCysIleSerSerCysCysArgProSerCysCysIleSerCysCysArgProSer 98
Db 331 TGTGTGCGCAGTCTGTGCTGCCAGCGGCTGTGTGACCTCTCTCCCTCCCTGCCAGCGGCC 390
Qy 99 CysCysArgThrThrCysCysArgProSerCysCysIle-----Ser 112
Db 391 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 450
Qy 113 SerCysCysArgProSerCysCysIleSerSerCysCysArgProSerCysCysArgThr 132
Db 451 CT-TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 509
Qy 133 ThrCysCysArgPro-----SerCysCysIle 141
Db 510 GTCTGTGTGCAAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 569
Qy 142 SerSerCysCysArgProSerCysCysIleSerSerCysCysCysArgProSerCysCysGln 161
Db 570 CAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 629

US-09-874-062-3 (1-230) x US-10-029-386-25183 (1-1002)
Qy 6 CysGlySerValCysSerAsp-----GlnSerCysGlyGlnGlyLeuGlyGln 21
Db 971 TGTGACTCT---TGCTCCGACTCTGTGCGAGTGTGAGCAGTGTG-----CCA 930
Qy 22 GluSerCysCysArgProSerCysCys----- 30
Db 929 GAGAGTGTGTGTGAGCCCTCTGTGCGCCCGCCCGCCCTGCTGAGCCTGTGTGTGTGCACC 870
Qy 31 -----GlnThrCysCysArgThrThrCys----- 39
Db 869 CCAGTGTGAGCGTGTGTGTCAGCCCTCTGTGCGCCAGTGTGTGAGCCAGCCCTGTGCAA 810
Qy 40 -----CysArgProSerCysCysIleSerSerCysCysArgProSer 53
Db 809 TCAGGCTGTGACAGCTCTGTGAGCCCTCTGTGCTGCCAGCAGTCTAGTGTGCCAGTGTGCT 750
Qy 54 CysCysIleSerSerCysCysLeuProSerCysCysLeuThrThrCysCysArgThrThr 73
Db 749 TGTGTGTGCTCTCTCCCTGTGCGAGCAGCCTGTGCTGTGCGCCGTGTGCTGTGCAAGACTGT 690
Qy 74 CysCysArgProSerCysCysIle-----SerSerCysCysArgPro 87
Db 689 TGTGTGCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 630
Qy 88 SerCysCysIleSerSerCysCysLeuProSerCysCysArgThrThrCysCysArgPro 107
Db 629 TCTAGTGTGCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 570
Qy 108 SerCysCys-----IleSerSerCysCysArgProSerCysCys 120
Db 569 GTCTGTGTGCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 510
Qy 121 IleSer-----SerCysCysLeuProSerCysCysArgThr 132
Db 509 GTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 450
Qy 133 ThrCys-----CysArgProSerCysCysIleSerSerCysCysArgProSerCys 149
Db 449 GGCTGTGACAGCTCTGTGACACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 390
Qy 150 CysIleSerSerCysCysLeuProSerCysCysGlnThrThrCysCysArgProSerCys 169
Db 389 TGCACCTCTCTCCCTGTGCGAGCAGCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 330
Qy 170 CysIleSerSerCysCysArg-----ProGlnCysCysGlnProSerCysCysArgPro 187
Db 329 TGTGTGCGCCAGCTGTGTGAGGATTCCTTCTTCTGCTGCCAGCAGTCTAGTGTGCCAGCA 270
Qy 188 AlaCysCysIleSerSerCysCysHisProSerCysCysValSerSerCysCysArgPro 207
Db 269 GTTGTGTGACCTCTCTCCCTGTGCGAGCAGCAGCCTGTGTGTGTGTGTGTGTGTGTGTGT 216
Qy 208 PheSerCysProThrThrCysCysArgThrThrCysPheHisProIleCysCysGlySer 227
Db 215 ---GGGGTTCACATCATGTGTGCGAGCAGTCTAGTGTGCCAGCGGCTGTGTGTGTGTGTGT 159
Qy 228 SerCysCys 230
Db 158 TCCTGTGTGC 150

RESULT 10
US-10-027-632-284090
; Sequence 284090, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
```


QY 162 ThrThrCysArgPro-----SerCysCys----- 170
DB 630 CCCATCTGCTGCAAGCCTGTCTGCTCTGGGATTCTCTTCTGCTGCCAGCAGTCTAGC 689
QY 171 IleSerSerCysTyArgProGlnCysCysGlnProSerCys----- 184
DB 690 TGTGTGAGCTGTGTCTCCAGCCCTGCTGCCAGCGGTCTGTGAGCCAGCCCTGCCAA 749
QY 185 -----CysArgProAlaCysCysIleSerSerCysCysHisProSer 198
DB 750 TCAGGCTGCATCAGCTCTGCTGCTGCTGCCAGCAGTCTAGCTGCMAGCCGGCT 809
QY 199 CysCysValSerSerCysArgCysProPheSerCysProThrThrCysCysArgThrThr 218
DB 810 TGTGTGACCTCTCC-----CCCTGCCAGCAGCGCTGTGTGCTGTG 854
QY 219 CysPheHisProIleCysCysGlySerSerCys 229
DB 855 TGTGTGCAAGCCCTGTCTGTGTGCCACCTGC 887

RESULT 11

US-10-027-632-284090
; Sequence 284090, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 284090
; LENGTH: 898
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-284090

Alignment Scores:

Pred. No.:	5,48e-41	Length:	898
Score:	615.50	Matches:	107
Percent Similarity:	48.11%	Conservative:	33
Best Local Similarity:	36.77%	Mismatches:	81
Query Match:	41.48%	Indels:	71
DB:	14	Gaps:	10

US-09-874-062-3 (1-230) x US-10-027-632-284090 (1-898)

QY 4 SerCysCysGlySerValCysArgAspInserCysGlyGlnGlyLeu----- 19
DB 31 GCCTGCTGCATGCCCGTCTGCTGCAAGACTGTCTGCTGCCAGCCTGTGTACTGTGTCCT 90
QY 20 -----GlyGlnGluSerCysArgProSerCysGlnThrThrCysCysArg 36
DB 91 GTCTGCAGTGGGGATTCTTATGCTGCCAGCAGTCTAGCTGCCAGTCTGTGCACC 150
QY 37 ThrThrCysCysArgProSerCysCysIleSerSerCysCysArgProSerCys----- 54
DB 151 TCCTCCCTCCAGCAGCGCTGTGTGTGCCATCTGTGTGCCAGCCTGTCTGTCTGTGGG 210

QY 55 CysIleSerSerCysCysLysProSerCysCysLeu-----ThrThrCysCys 70
DB 211 ATTTCTCTTCTGCTGCCAGCTAGCTGTGTGAGCTGTGTGTCAGYCCCTGTCTGC 270
QY 71 ArgThrThrCys-----CysArgProSer 78
DB 271 CAGGCGGTCTGTGAGCCAGCCCTGCCAATCAGCTGCATCAGCTCCTGCACGCCCTCG 330
QY 79 CysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCysLysProSer 98
DB 331 TGTCTGCCAGCAGTCTAGCTGCCAGCGGTCTGCTGCACCTCTCTCCCTGCCAGCAGCC 390
QY 99 CysCysArgThrThrCysCysArgProSerCysCysIle-----Ser 112
DB 391 TGTCTGTGCTGTCTGCTGCCAGCCGTCTCTGTGTGCCACCTCTCTCTGATGATTCG 450
QY 113 SerCysCysArgProSerCysCysIleSerSerCysCysLysProSerCysCysArgThr 132
DB 451 CT-TGCTGTGCTCTCTCCCTGCCAGCAGCGCTGTGCATGCCGTCTGTCTGCAAGACT 509
QY 133 ThrCysCysArgPro-----SerCysCysIle 141
DB 510 GTCTGCTGCCAGCCTGTGTCTGTGCTGTCTGCTGGGATTCTTTCATGCTGCCAG 569
QY 142 SerSerCysCysArgProSerCysCysIleSerSerCysCysLysProSerCysCysGln 161
DB 570 CAGTCTAGCTGCCAGTCTGCTGCCAGCTCTCTCCCTGCCAGCAGCGCTGTGTGTG 629
QY 162 ThrThrCysCysArgPro-----SerCysCys----- 170
DB 630 CCCATCTGCTGCAAGCCTGTCTGCTCTGGGATTCTCTCTGCTGCCAGCAGTCTAGC 689
QY 171 IleSerSerCysTyArgProGlnCysCysGlnProSerCys----- 184
DB 690 TGTGTGAGCTGTGTGTGCCAGCCCTGCTGCCAGCGGTCTGTGAGCCAGCCCTGCCAA 749
QY 185 -----CysArgProAlaCysCysIleSerSerCysCysHisProSer 198
DB 750 TCAGGCTGCATCAGCTCTGCTGCCAGCCCTGCTGCTGCCAGCAGTCTAGCTGCMAGCCGGCT 809
QY 199 CysCysValSerSerCysArgCysProPheSerCysProThrThrCysCysArgThrThr 218
DB 810 TGTGTGACCTCTCTCC-----CCCTGCCAGCAGCGCTGTGTGCTGTGCTGTG 854
QY 219 CysPheHisProIleCysCysGlySerSerCys 229
DB 855 TGTGTGCAAGCCCTGTCTGTGTGCCACCTGC 887

RESULT 12

US-09-950-051-7/c
; Sequence 7, Application US/09950051
; Publication No. US20030129594A1
; GENERAL INFORMATION:

; APPLICANT: VERAMIAN, EDWARD
; TITLE OF INVENTION: GENES AND THE PHYSICS OF THE DNA DOUBLE HELIX.
; TITLE OF INVENTION: FORMULATION OF A PHYSICS-BASED GENE IDENTIFICATION
; TITLE OF INVENTION: (PBTI) METHOD: AB INITIO IDENTIFICATION OF GENES IN
; TITLE OF INVENTION: EUKARYOTIC GENOMES
; FILE REFERENCE: 03495-0209-00000
; CURRENT APPLICATION NUMBER: US/09/950,051
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 60/232,146
; PRIOR FILING DATE: 2000-09-13
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1800
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-950-051-7

Alignment Scores:


```
Pred. No.: 1.19e-39 Length: 1800
Score: 603.00 Matches: 108
Percent Similarity: 56.22% Conservatives: 23
Best Local Similarity: 46.35% Mismatches: 68
Query Match: 40.63% Indels: 34
DB: 13 Gaps: 8

US-09-874-062-3 (1-230) x US-09-950-051-7 (1-1800)

Qy 25 CysArgProSerCysCysGlnThrThrCys-----CysArgThrThrCysCysArg 41
Db 1795 TGTGAGCCAGCCCTGTCATCAGCTGCACAGCTCTCTGAGCCCTCTGCTGTCAG 1736

Qy 42 ProSerCysCysLeuThrThrCysCysArgProSerCysCysLeuThrThrCys 61
Db 1735 CCGCTTCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1676

Qy 62 ProSerCysCysLeuThrThr-----CysCysArgThrThrCys 74
Db 1675 CTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1616

Qy 75 CysArgProSerCysCysLeuThrThrCysCysArgProSerCysCysLeuThrThrCys 94
Db 1615 TCCAGCCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1556

Qy 95 CysLeuProSerCysCysArgThrThrCysCysArg-----ProSerCysCysLeuThr 112
Db 1555 TCCAGCCCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1496

Qy 113 SerCysCysArgProSerCysCysLeuThrThrCysCysArgProSerCysCysArgThr 132
Db 1495 TCTAGTCCAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1436

Qy 133 ThrCysCysArgProSerCysCysLeuThrThrCysCysArgPro----- 147
Db 1435 GTCTGTTCAAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1376

Qy 148 -----SerCysCysLeuThrThrCysCysArgProSerCysCysGlnThrCysCys 165
Db 1375 TCCACTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1316

Qy 166 ArgProSerCysCysLeuThrThrCysCysArgProGlnCysCysGlnProSerCysCys 185
Db 1315 AGACCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1274

Qy 186 ArgProAlaCysCys-----IleSerSerCysCysHisPro---SerCysCysValSer 202
Db 1273 AGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1214

Qy 203 SerCysCysArgProSerCysCysArgProThrThrCysCysArgThrThrCysCysPheHisPro 222
Db 1213 AGCTGCTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1154

Qy 223 IleCysCysGly-----SerSerCys 229
Db 1153 GCCTGCTGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1115

RESULT 13
US-09-864-761-6314/c
; Sequence 6314, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6314
; LENGTH: 478
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006070.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
; US-09-864-761-6314

Alignment Scores:
Pred. No.: 3.03e-40 Length: 478
Score: 602.50 Matches: 92
Percent Similarity: 65.79% Conservatives: 33
Best Local Similarity: 48.42% Mismatches: 34
Query Match: 40.60% Indels: 31
DB: 9 Gaps: 6

US-09-874-062-3 (1-230) x US-09-864-761-6314 (1-478)

Qy 28 SerCysCysGlnThrThrCysCysArgThrThrCysCysArgProSerCysCysLeuThr 47
Db 477 ACATGCTGCTGAGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 430

Qy 48 SerCysCysArgProSerCysCysLeuThrThrCysCysArgProSerCysCysLeuThr 67
Db 429 -----ACTGTGACC 421

Qy 68 ThrCysCysArgThrThrCysCysArgProSerCysCysLeuThrThrCysCysArgPro 87
Db 420 ACCTGACGACGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 361
```

```
Qy 88 SerCysCysLysSerSerCysCysLysProSerCysCysArgThrThrCysCysArgPro 107
Db 360 ---TGCTGCCACCACTTGTCTGCAAAACACCTGCTGTCAGGACCACTGCTGCCAGCCC 304
Qy 108 SerCysCysLysSerSerCysCysArgProSerCysCysLysSerSerCysCysLysPro 127
Db 303 ACT---TGTGTGGCAGCTGCTGCCAGCCTTCTGCTGCAGACACACCTGCTGCCAGCCC 247
Qy 128 SerCysCysArgThrThrCysCysArgProSerCysCysLysSerSerCysCysArgPro 147
Db 246 ACTTGCTGTGGTCCAGCTGCTGTCGCAACACAGCTGTGGTCCAGCTGCTGTCAGCCT 187
Qy 148 SerCysCysLysSerSerCysCysLysProSerCysCysGlnThrThrCysCysArgPro 167
Db 186 ---ATTGTGGGTCCAGTTGCTGTCAGCCT-----TGTGTTCACCCG 148
Qy 168 SerCysCysLysSerSerCysCysArgProGlnCysCysGlnProSerCysCysArgPro 187
Db 147 ACTTGCTATCAAACTATCTGCTTCAGGACCACTGCTGCCAGCCTACTGCTGCCAGCCC 88
Qy 188 AlaCysCysLysSerSerCysCysHisProSerCysCysValSerSerCysCysArgPro 207
Db 87 ACTGCTGCAGAACACCTCTTGCAGCCCACTGCTGTCGGTTCAGCTGC---TGCCAG 31
Qy 208 PheSerCysProThrThrCysCysArgThr 217
Db 30 CTTGCTGCCACCAACATGCTGTCAAAAC 1
RESULT 14
US-10-029-386-24945/c
; Sequence 24945, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24945
; LENGTH: 865
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP001067.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.25
; OTHER INFORMATION: EST HUMAN HIT: BF057518.1, EVALUE 4.00e-62
; OTHER INFORMATION: SWISSPROT HIT: O75690, EVALUE 1.00e-04
; OTHER INFORMATION: NT HIT: AL163300.2, EVALUE 0.00e+00
US-10-029-386-24945
Alignment Scores:
Pred. No.: 7,67e-40 Length: 865
Score: 601.00 Matches: 101
Percent Similarity: 59.91% Conservative: 29
Best Local Similarity: 46.54% Mismatches: 69
Query Match: 40.50% Indels: 18
DB: 13 Gaps: 7
US-09-874-062-3 (1-230) x US-10-029-386-24945 (1-865)
Qy 23 SerCysCysArgProSerCysCysGlnThrThrCysCysArgThrThrCysCysArgPro 42
Db 857 AGCTCTGCAGCGCCTCGTGTGTCGACCGGCTGTCGCGCTCTCTCCCTGCCAGCAG 798
Qy 43 SerCysCysLysSerSerCysCysArgProSerCysCysLysSerSerCysCysLysPro 62
Db 797 GCCTGCTGCTGCTGCTGCTGCAAGCCTGTCGCTGCCCACTGCTCTAAGGAT 738
```

```
Qy 63 SerCysCysLeuThrThrCysCysArgThrThrCysCysArgProSerCysCysLysSer 82
Db 737 TCC-----TCCTCATGCTGCCAGCAGTCTAGCTGCCAGCAACTTGTGTGCTCT 687
Qy 83 SerCysCysArgProSerCysCysLysSerSerCysCysLysProSerCysCysArgThr 102
Db 686 TCTCTCTGCAGCAGTCTGCTGTGTCCTGCTGTCGCAAGCCCGCTGTGTGTGCCCC 627
Qy 103 ThrCysCysArg-----ProSerCysCysLysSerSerCysCysArgProSerCysCys 120
Db 626 ACCTGTTCTGAGGATTCCTCTTCATGCTGCCAGCATCTAGCTGCCAGCGACTGCTGC 567
Qy 121 IleSerSerCysCysLysProSerCysCysArgThrThrCysCysArgProSerCysCys 140
Db 566 ACCTCTCCCTGCCAGCAGTCTCTAGTCCCTGCTGTCGTCCTGTCGTCGTCGTCGTC 507
Qy 141 IleSerSerCysCysArgPro-----SerCysCysLysSerSer 153
Db 506 AACCCTATCTGCTGTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 447
Qy 154 CysCysLysProSerCysCysGlnThrThrCysCysArgProSerCysCysLysSerSer 173
Db 446 AGCTGTGCAGCGCTGCTGTCACCACTCTGCTGTCGACAGCCCTCTCTCCGTCGTCCTC 387
Qy 174 CysTyArgProGlnCysCysGlnProSerCysCysArgProAlaCysCysLysSerSer 193
Db 386 CTCTGCCGCC---ATATGCAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 339
Qy 194 CysCysHisPro---SerCysCysValSerSerCysCysArgProPheSerCysProThr 212
Db 338 TGTGTGCCCTGCTGCTCTCTACCAGCGCAGTGTGCTGCCCGCGCTCTCTGCTGCTGCC 279
Qy 213 ThrCysCysArgThrThrCysCysPheHisProIleCysCysGlySerSerCys 229
Db 278 CTCCTCTGCCCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 231
RESULT 15
US-09-864-761-4747
; Sequence 4747, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
```

;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 4747
;; LENGTH: 1974
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC003958.1
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.7
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2
;; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.3
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
US-09-864-761-4747

Alignment Scores:

Pred. No.:	1.9e-39	Length:	1974
Score:	601.00	Matches:	112
Percent Similarity:	45.86%	Conservative:	32
Best Local Similarity:	35.67%	Mismatches:	76
Query Match:	40.50%	Indels:	94
DB:	9	Gaps:	10

US-09-874-062-3 (1-230) x US-09-864-761-4747 (1-1974)

QY	2	ValSerSerCysCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGlyGln	21
DB	15	GTGAGTAGCTGTGCCCAACCCCTGTCTGTGAT	47
QY	22	GlusSerCysCysArgProSerCys	36
DB	48	CCTGTCAATTGTGAGCCTTCTTGCTCCGTGAGCAGCGGCTGCCAACCCGTGTGTGAG	107
QY	37	ThrThrCysCysArgProSerCysCysIleSerCysCysArgProSerCysCysIle	56
DB	108	GCACCAACCTGTGAGCCTTCTTGCTCTGTGAGCAACTGTCTACCAACCTGTGTCTCGAG	167
QY	57	SerSerCysCysLysProSerCysCysLeuThrThrCysCysArgThrThrCys	74
DB	168	GCACCATCTGTGAGCCTTCTTGCTCAGTGAGCAACTGTCTGCCAACCTGTGTGCTTTGAG	227
QY	75	-----CysArgProSerCysCysIleSerSerCys	84
DB	228	GCACCGTTTGTGAGCCTTCTTGCTCAGTGAGCAACTGTCTGCCAACCTGTGTGCTGTGAG	287
QY	85	-----CysArgProSerCysCysIleSerSerCysCysLysPro	97
DB	288	CTGTGATTGTGAGCCTTCTTGCTCCGTGAGCAGCTGTCTGCCAGCCTGTAGGCTCTGAA	347
QY	97	-----	97
DB	348	GCACATTCCTGCCAACAGTCCTCTGTGTGCCCACTTCTGCCAGGCTGTCTCTGCAAA	407
QY	98	---SerCysCysArgThrThrCysCysCysArgProSerCysCysIleSerSerCys	114

Search completed: December 13, 2003, 07:36:57

Job time : 338 secs

DB	408	TCCAGCTGCTGCCAGCCAGTTGTCTGTGAGCCAGCTGTCTGTTCAGCTGTCTGCAACCTG	467
QY	115	-----CysArgProSerCysCys	120
DB	468	CCTAGTTCTCTGCCAACCTGTGTGTGAGCCTTCTCTGTCTGTGAGCGGTGTGCCGACA	527
QY	121	-----IleSerSerCys	126
DB	528	CCTACCTGCTCTGTGACCACTGTGTGTGAGCCAGCTGTCTGTGACCCAGCCCTTGTGAG	587
QY	127	ProSerCysCysArgThrThrCysCysArgProSerCysCysIleSer	142
DB	588	CCAGTTCTCAGAGTCTAGCATCTGCCAGCAGCTAGTGTGTGTGCTGTGTGCTGTGAG	647
QY	143	---SerCysCysArgProSerCysCysIleSerSerCysCysLysProSerCysCysGln	161
DB	648	CCAGTTTGCCTCCGCCCTGT	707
QY	162	ThrThrCysCysArgProSerCysCysIleSerSerCysCysTyrArgProGlnCysCysGln	181
DB	708	AGCACTTGCCAAAGAGCCTTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	767
QY	182	ProSerCysCysArgProAlaCysCysIleSerSerCysCysHisProSerCys	199
DB	768	CCAGCCCTCTCTACCACTGT	827
QY	200	-----CysValSerSerCysArgCysProPheSerCysProThrThrCysCysArg	216
DB	828	GTCAAGCGCTGTCTCTGT	887
QY	217	ThrThrCysPheHisProIleCysCysGlySerSerCysCys	230
DB	888	CCTCTTTCTGTGAGTCCAGGGTCTTCTGTGCATCTGCCATCTGC	929

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 13, 2003, 05:43:14 ; Search time 67 seconds

(without alignments)
1515.196 Million cell updates/sec

Title: US-09-874-062-3

Perfect score: 1484

Sequence: 1 MWSSCCGVSCDQSGGGLG.....PTTCRTTCFHPICGSSCC 230

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09874062/runat 12122003 171323 24954/app_query.fasta_1.391
-DB=Issued Patents NA -QFMT=fastcap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NORW=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09874062 @CGN 1 1 56 @runat 12122003 171323 24954 -NCPUS=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	733	49.4	3489	2	US-08-728-323A-1
C 2	733	49.4	3489	4	US-09-298-568-1
C 3	733	49.4	3489	4	US-09-410-399-1
C 4	733	49.4	32207	2	US-08-770-379-20
C 5	733	49.4	32207	3	US-08-757-669A-20
C 6	733	49.4	32207	4	US-09-230-371A-20
C 7	445	30.0	1926	4	US-09-249-588A-2
C 8	445	30.0	1926	4	US-09-410-399-3
C 9	445	30.0	2580	3	US-09-050-863-2
C 10	445	30.0	2580	4	US-09-359-081-2
C 11	445	30.0	5452	2	US-09-130-114-1
C 12	445	30.0	8705	4	US-09-647-344A-14

C 13	445	30.0	9600	3	US-08-910-647-1
C 14	445	30.0	9600	4	US-09-620-925-1
C 15	445	30.0	10596	1	US-07-884-811-15
C 16	445	30.0	10596	1	US-07-885-971-15
C 17	445	30.0	10596	1	US-08-087-783A-15
C 18	445	30.0	10596	1	US-08-194-088B-15
C 19	445	30.0	10596	2	US-08-194-087-15
C 20	445	30.0	10596	5	PCT-US93-04648-15
C 21	429.5	28.9	9551	1	US-08-056-200-93
C 22	429.5	28.9	9551	2	US-08-800-644-93
C 23	398.5	26.9	5163	3	US-08-700-651-1
C 24	398.5	26.9	5163	3	US-08-928-361B-4
C 25	398.5	26.9	5163	4	US-09-588-995A-4
C 26	398.5	26.9	5318	3	US-08-700-651-2
C 27	398.5	26.9	5318	3	US-08-928-361B-3
C 28	398.5	26.9	5318	4	US-09-588-995A-3
C 29	381	25.7	1562	4	US-09-620-312D-510
C 30	354.5	23.9	543	6	5273901-6
C 31	341	23.0	688	3	US-08-998-416-915
C 32	340.5	22.9	4403765	3	US-09-103-840A-2
C 33	340.5	22.9	4411529	3	US-09-103-840A-1
C 34	336.5	22.7	5511	3	US-08-928-361B-2
C 35	336.5	22.7	5511	4	US-09-588-995A-2
C 36	336.5	22.7	7334	3	US-08-928-361B-1
C 37	336.5	22.7	7334	4	US-09-588-995A-1
C 38	335.5	22.6	4403765	3	US-09-103-840A-2
C 39	335.5	22.6	4411529	3	US-09-103-840A-1
C 40	334.5	22.5	1107	2	US-08-991-300-1
C 41	330.5	22.3	533	6	5482709-5
C 42	325.5	21.9	2296	2	US-08-899-336-2
C 43	323	21.8	6558	4	US-09-491-356C-7
C 44	307.5	20.7	4262	4	US-09-521-511C-10
C 45	301.5	20.3	4350	4	US-09-295-593-37

ALIGNMENTS

RESULT 1

US-08-728-323A-1/C
; Patent 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728.323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400


```

Db      1571  GGCTCCTGCTGCTGGCTCCTGCTGTGTGGCTTCCTGCTGTGGCTCTGCTGTTGT 1511
Qy      212   ThrThrCysCysArgThrThrCysPheHisProIleCysCysGlySerSerCysCys 230
           ::::::::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
Db      1511  GGCTCCTGCTGTTGTGGCTCCTGCAGGGCTCCTGCTGCTGTGGCTCCCTGCTGTTGT 1455

RESULT 3
US-09-410-399-1/c
; Sequence 1, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Cotter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; TITLE OF INVENTION: to Genomic Host DNA
; FILE REFERENCE: UM-03778
; CURRENT APPLICATION NUMBER: US/09/410,399
; CURRENT FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-1

Alignment Scores:
Pred. No.:          3,12e-42              Length:          3489
Score:             733.00                 Matches:         122
Percent Similarity: 57.32%                Conservative:     15
Best Local Similarity: 51.05%               Mismatches:       88
Query Match:        49.39%                  Indels:          14
DB:                 4                      Gaps:            4

US-09-874-062-3 (1-230) x US-09-410-399-1 (1-3489)

Qy      3   SerSerCysCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGlyGlnGlu 22
Db      2162 TCCTGCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCA 2112
           |||||
Qy      23   SerCysCysArg---ProSerCysCysGlnThrThrCysCysArgGthrThrCysCysArg 41
           |||||
Db      2111 TCCTGCTGCTTCCCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCATCTGCTGCTGC 2052
           |||||
Qy      42   ProSerCysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCysLys 61
           |||||
Db      2051 TCATCCTGCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGC 1992
           |||||
Qy      62   ProSerCysCysLeuThrThrCysCysArgGthrThrCysCysArgProSerCysCysIle 81
           |||||
Db      1991 TCATCCTGCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGC 1932
           |||||
Qy      82   SerSerCysCysArgProSerCysCysIleSerSerCysCysLysProSerCysCysArg 101
           |||||
Db      1931 TCATCCTGCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGC 1872
           |||||
Qy      102  ThrThrCysCysArgProSerCysCysIleSerSerCysCysArgProSerCysCysIle 121
           |||||
Db      1871 TCATCCTGCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGC 1812
           |||||
Qy      122  SerSerCysCysLysProSer---CysCysArgGthrThrCysCysArgProSerCysCys 140
           |||||
Db      1811 TCATCCTGCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGT 1752
           |||||
Qy      141  IleSerSerCysCysArgProSerCysCysIleSerSerCysCysLysProSerCysCys 160
           |||||
Db      1751 GGCTCCCCTGCTGCTGGCTCCCGCTGCTGCTGGCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGT 1692
           |||||
Qy      161  GlnThrThrCys-----CysAArgProSerCysCysIle 171
           :::|
Db      1691 GGCTCCCCTGCTGGGGCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGT 1632
           |||||
Qy      172  SerSerCysbYArqPrroGlnCysCysGlnProSerCysCysArgPrroAlaCysCysIle 191
           |||||

```


[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 13, 2003, 04:44:18 ; Search time 283 Seconds
(without alignments)
2193.890 Million cell updates/sec

Title: US-09-874-062-3

Perfect score: 1484

Sequence: 1 MVSSCGSVCSQSCQGLG.....PTTCRTTCFHPICGSSCC 230

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=xlh
-Q=/cgn2_1/USPTO_spool/US09874062/runat_12122003_171322_24924/app_query.fasta_1.391
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09874062 @CGN_1_1_0 @runat_12122003_171322_24924 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPEXT=7
-FGAPOPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq 19Jun03: **
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT: *
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT: *
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT: *
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT: *
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT: *
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT: *
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT: *
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT: *
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT: *
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT: *
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT: *
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT: *
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT: *
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT: *
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT: *
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT: *
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT: *
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT: *
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT: *
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT: *
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT: *
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: *
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT: *
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT: *
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
1	1484	100.0	1101	22	AAD06303	Human cDNA clone H	
2	1229.5	82.9	877	24	ABZ11932	Human polynucleoti	
3	1218.5	82.1	627	23	AAS93555	DNA encoding novel	
4	1053	71.0	911	25	ABX70816	Novel human cDNA s	
C	5	944.5	63.6	1851	22	AAK52135	Human polynucleoti
6	944.5	63.6	1851	22	AAK53119	Human polynucleoti	
7	767	51.7	1851	22	AAK52135	Human polynucleoti	
C	8	767	51.7	1851	22	AAK53119	Human polynucleoti
9	733	43.4	3489	21	AAA30290	Kaposi's sarcoma-a	
C	10	733	49.4	3489	22	AAF82901	Nucleotide sequenc
C	11	733	49.4	3489	24	ABA93487	Kaposi's sarcoma-a
12	733	49.4	32207	20	AAV73805	KSHV LUR DNA (nucl	
13	733	49.4	137507	19	AAV19941	KSHV long unique c	
14	666.5	44.9	1980	24	ABK48978	Physics-based gene	
15	603	40.6	410	22	ABA09521	Human hair keratin	
C	16	603	40.6	1800	24	ABK48979	Physics-based gene
C	17	602.5	40.6	478	22	ABA59018	Human foetal liver
C	18	602.5	40.6	478	22	ABA27848	Probe #6314 for ge
C	19	602.5	40.6	478	22	AAK07188	Human brain expres
C	20	602.5	40.6	478	22	AAK32938	Human bone marrow
C	21	602.5	40.6	478	22	AAI16266	Probe #6199 for ge
C	22	602.5	40.6	478	22	AAI38744	Probe #7430 used t
C	23	602.5	40.6	478	23	ABS32666	Human liver single
C	24	602.5	40.6	478	24	ABS07743	Human genome-deriv
25	601	40.5	1974	22	ABA46130	Human breast cell	
26	601	40.5	1974	22	ABA56676	Human foetal liver	
27	601	40.5	1974	22	ABA26281	Probe #4747 for ge	
28	601	40.5	1974	22	AAK04812	Human brain expres	
29	601	40.5	1974	22	AAK30335	Human bone marrow	
30	601	40.5	1974	22	AAI14939	Probe #4872 for ge	
31	601	40.5	1974	22	AAI36293	Probe #4979 used t	
32	601	40.5	1974	22	AAI04716	Probe #4707 used t	
33	601	40.5	1974	23	ABS29975	Human liver single	
34	601	40.5	1974	24	ABS04922	Human genome-deriv	
35	600.5	40.5	1067	24	ABQ61190	Keratin associated	
36	574	38.7	489	22	ABA70389	Human foetal liver	
37	574	38.7	489	22	AAK18627	Human brain expres	
38	574	38.7	489	22	AAK44553	Human bone marrow	
39	574	38.7	489	22	AAI50538	Probe #19224 used	
40	574	38.7	489	23	ABS44211	Human liver single	
41	574	38.7	489	24	ABS18790	Human genome-deriv	
42	570.5	38.4	461	22	ABA57772	Human foetal liver	
43	570.5	38.4	461	22	AAK05838	Human brain expres	
44	570.5	38.4	461	22	AAK31470	Human bone marrow	
45	570.5	38.4	461	22	AAI37355	Probe #6041 used t	

ALIGNMENTS

RESULT 1

AAD06303

ID AAD06303 standard; cDNA; 1101 BP.

XX AAD06303;

AC

DT 10-AUG-2001 (first entry)

DE Human cDNA clone HHQC55 encoding FDCD-containing protein.

XX Human; four disulfide core domain; FDCD; immunosuppressive; cytostatic;
KW antarthritic; antirheumatic; antiproliferative; cardiant; vasotrophic;
KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; dermatological; antinflammatory; vulnary;
KW antarteriosclerotic; gene therapy; skin disorder; congenital disorder;
KW mole; freckle; haemangioma; integumentary tumour; basal cell carcinoma;
KW keratosis; melanoma; atherosclerosis; urticaria; photosensitivity;

CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 627 BP; 89 A; 240 C; 151 G; 147 T; 0 other;

Alignment Scores:

Pred. No.: 4,45e-48 Length: 627
 Score: 1218.50 Matches: 194
 Percent Similarity: 84.35% Conservative: 0
 Best Local Similarity: 84.35% Mismatches: 1
 Query Match: 82.11% Indels: 35
 DB: 23 Gaps: 1

US-09-874-062-3 (1-230) x AAS93555 (1-627)

QY 1 MetValSerSerCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGly 20
 DB 36 ATGGTCAGCTCTGCTGTGGCTCTGCTGTCTGTGACAGAGCTGTGGTCAAGGCTCTCGC 95
 QY 21 GlnGluSerCysCysArgProSerCysCysGlnThrThrCysCysArgThrThrCysCys 40
 DB 96 CAGGAGAGCTGTGCGCGCCAGCTGTGCTCCAG----- 128
 QY 41 ArgProSerCysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCys 60
 DB 128 ----- 128
 QY 61 LysProSerCysCysLeuThrThrCysCysArgThrThrCysCysArgProSerCysCys 80
 DB 129 -----ACCACTGCTGCAGGACCACTGCTGCGCCAGCTGTGC 170
 QY 81 IleSerSerCysCysArgProSerCysCysIleSerSerCysCysIleSerProSerCysCys 100
 DB 171 ATTTCCAGTTGTGTCAGGCTCTCTGCTGTATCTCCAGCTGTGCACACCAAGCTGTGC 230
 QY 101 ArgThrThrCysCysArgProSerCysCysIleSerSerCysCysArgProSerCysCys 120
 DB 231 AGGACCACCTGCTGCGCGCCAGCTGTGCTGCTATTTCCAGTTGCTTCCAGGCTTCTGCTGT 290
 QY 121 IleSerSerCysCysIleSerProSerCysCysArgThrThrCysCysArgProSerCysCys 140
 DB 291 ATCTCCAGCTGTGCAACCCAGCTGTGCTGAGACCACTGCTGCGCCAGCTGTGC 350
 QY 141 IleSerSerCysCysArgProSerCysCysIleSerSerCysCysIleSerProSerCysCys 160
 DB 351 ATTTCTAGTTGTGTCAGGCTTCTCTGCTGTATCTTAGCTGTGCACCAAGCTGTGC 410
 QY 161 GlnThrThrCysCysArgProSerCysCysIleSerSerCysCysIleSerProGlnCysCys 180
 DB 411 CAGACCACCTGCTGCGCGCCAGCTGTGCTATCTCCAGCTGTGTACAGGCCCCAGTGTGC 470
 QY 181 GlnProSerCysCysArgProAlaCysCysIleSerSerCysCysHisProSerCysCys 200
 DB 471 CAGCCCTCTGCTGCGCGCCGCTGCTGCTATTTCTAGTTGTGTGTCATCCAGCTGTGT 530
 QY 201 ValSerSerCysArgCysProPheSerCysProThrThrCysCysArgThrThrCysPhe 220
 DB 531 GTGTCCAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 590
 QY 221 HisProIleCysCysGlySerSerCysCys 230
 DB 591 CACCCCATCTGCTGCGGAGTTCTTCTGCTGC 620

RESULT 4

ABX70816
 ID ABX70816 standard; cDNA; 911 BP.
 XX
 AC ABX70816;
 XX
 DT 05-MAR-2003 (first entry)
 XX
 DB Novel human cDNA sequence #41.
 XX
 KW Human; gene; ss; nervous system disorder; peripheral neuropathy;
 KW Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
 KW neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
 KW autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
 KW insulin-dependent diabetes mellitus; anaemia; thrombocytopenia; wound;
 KW ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
 KW fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
 KW coagulation disorder; cancer; tumour; inflammatory disease;
 KW septic shock; Crohn's disease; anaphylaxis; proliferation; chemotactic;
 KW differentiation; stem cell growth factor; haematopoiesis; chemokinetic;
 KW haemostatic; antiinflammatory; expressed sequence tag; EST.
 XX
 OS Homo sapiens.
 XX
 PN WO200281731-A2.
 XX
 PD 17-OCT-2002.
 XX
 PP 29-JAN-2002; 2002MO-US01222.
 XX
 PR 30-JAN-2001; 2001US-0774528.
 XX
 PA (HYSE-) HYSEO INC.
 PA (GOOD/) GOODRICH R W.
 XX
 PI Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;
 XX
 DR WPI, 2003-058563/05.
 XX
 PT Novel polypeptide useful for treating neurodegenerative diseases,
 PT myeloid or lymphoid cell disorders, bone disorders, mechanical and
 PT traumatic disorders, coagulation disorders, and inflammatory diseases
 PT -
 XX
 PS Claim 1; Page -: 612pp; English.
 XX
 CC This invention relates to the cDNA sequences encoding an isolated
 CC novel human polypeptide. The protein encoded by the nucleic acid of
 CC the invention is useful for treating central and peripheral nervous
 CC system diseases (e.g. peripheral neuropathy, Huntington's disease,
 CC amyotrophic lateral sclerosis); neurodegenerative diseases (e.g.
 CC Parkinson's disease, Alzheimer's disease); autoimmune disease (e.g.
 CC systemic lupus erythematosus, rheumatoid arthritis, insulin-dependent
 CC diabetes mellitus); myeloid or lymphoid cell disorders (e.g. anaemia
 CC and thrombocytopenia); wounds, ulcers, burns; bone disorders (e.g.
 CC osteoporosis, osteoarthritis); mechanical and traumatic disorders (e.g.
 CC stroke, head trauma); lung or liver fibrosis; reperfusion injury in
 CC various tissues; bacterial, viral or fungal infections; allergic
 CC conditions such as allergic rhinitis, asthma; coagulation disorders
 CC (e.g. haemophilia); cancer and tumours; and inflammatory diseases (e.g.
 CC septic shock, Crohn's disease, anaphylaxis). The protein may be used to
 CC inhibit the growth, infection or function of infectious agents such as
 CC bacteria, fungi, viruses, or to effect bodily characteristics,
 CC biohythms or circadian cycles of rhythms. The protein may also
 CC have proliferative/differentiation, immune stimulating or suppressive,
 CC haematopoietic regulation, haemostatic and thrombolytic, receptor/ligand,
 CC chemotactic/chemokinetic, and an antiinflammatory activities. The cDNA sequences of the invention are
 CC useful for expressing recombinant protein for analysis. The present
 CC sequence represents a novel human cDNA sequence of the invention,
 CC this sequence is an expressed sequence tag (EST) and was identified
 CC using subtractive hybridisation.


```
QY 41 ArgProSerCysCysIleSerSerCysCysArgProSer-----CysCys 55
Db 805 ---CCAGCTGTGTATCCAGCTGCTGCCGCCATCTCTCTCAGACTACTGCTGTC 749
QY 56 IleSerSerCysCysLysProSerCysCysLeuThrThrCysCysArgThrThrCysCys 75
Db 748 CAGACCACTTGTGTGCGCCCGAGCTGCTGCCACCCAGTCTGTGTGTCAGACCACC---TGC 692
QY 76 ArgProSerCysCysIleSerSerCysCysArgProSerCysCys----- 90
Db 691 CGCCCCAGCTGTGTGTCTCCAGCTGCTGCCGCCACCTCTCTGTCAGACCACCTGCCAC 632
QY 91 -----IleSerSerCysCysLysProSerCysCysArgThrThrCysCysArg 106
Db 631 CCCAGCTGTGTATGTCAGCTGCTGCCGTCCTCACTCTGTGTGTCAGACCACC---TGC 575
QY 107 ProSerCysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCysLys 126
Db 574 CCCAGCTGTGTGTCTCCAGCTGCTGCCGTCCTCACTCTGTGTGTCAGACCACCTGCTGCCGT 515
QY 127 ProSerCysCysArgThrThr----- 133
Db 514 GCAACTTGTGCCGCCCGCCAGCTGCTGTGTGATCTCTTGTGAACCTCATATTGGAATATC 455
QY 134 -----CysCysArgProSer----- 138
Db 454 AACCATGAGCCAGTCACCATCCATGATATGAAAAGAACTGCTGCCGTCCTCCAGCTGCTGCC 395
QY 139 -----CysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCys 155
Db 394 CAGACCACTTGTGTCAGGACCACTGCTGCCGCCCGCCAGCTGCTGTGTGTCAGCTGCTGCC 335
QY 156 LysProSerCysCysGlnThrThrCysCysArgProSerCysCysIleSerSerCysTyr 175
Db 334 AGACCGCAGTGTGTCAGTCTGTGTGCTGCCAGCCACCTGCTGTGAGCCCGCCAGCTGCTGCC 275
QY 176 ArgProGlnCysCysGlnProSerCysCysArgProAlaCysCysIleSerSerCysCys 195
Db 274 CAGACCACTTGTGTCAGGACCACTGCTGCCGTCCTCCAGCTGCTGTGTGTCAGCTGCTGCC 215
QY 196 HisProSerCysCysValSer-----SerCysArgCysProPheSerCys 210
Db 214 AGACCCCGAGTGTGTCAGCTCTGTGTACTGCCAGCCACCTGCTGCCGCCCGCCAGCTGCTGCC 155
QY 211 ProThrThrCysCysArgThrThrCysCysPheHisProLysCysGlySerSerCysCys 230
Db 154 CAGACCACTTGTGTCAGGACCACTGCTACCCGCCCGCCAGCTGCTGTGTGTCAGCTGCTGCC 95

RESULT 6
AAK53119
ID AAK53119 standard; cDNA; 1851 BP.
XX
AC AAK53119;
DT
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 2648.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
```

```
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT, Auundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX WPI; 2001-476283/51.
XX P-PSDB; AAM79986.
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX Claim 1; Page 4903; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM80020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.
XX SQ Sequence 1851 BP; 340 A; 574 C; 558 G; 379 T; 0 other;

Alignment Scores:
Pred. No.: 3,54e-35 Length: 1851
Score: 944.50 Matches: 153
Percent Similarity: 66.07% Conservative: 32
Best Local Similarity: 54.64% Mismatches: 38
Query Match: 63.65% Indels: 57
DB: 22 Gaps: 9

US-09-874-062-3 (1-230) x AAK53119 (1-1851)
QY 1 MetValSerSerCysCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGly 20
Db 939 ATGGTTAACTCTTGTGTGGCTCTGTCTCTGACCCAGGGCTGTGTATCAAGGCTCTGC 998
QY 21 GlnGluSerCysCysArgProSerCysCysGlnThrThrCysCysArgThrThrCysCys 40
Db 999 CAAGAGACCTGCTGCCGCCCGCCAGCTGCTGCCAGACCACCTGT-----TGCTGC 1046
QY 41 ArgProSerCysCysIleSerSerCysCysArgProSer-----CysCys 55
Db 1047 ---CCCAGCTGTGTGTATCCAGCTGCTGCCGCCATCTCTCTCAGACTACTGCTGTC 1103
QY 56 IleSerSerCysCysLysProSerCysCysLeuThrThrCysCysArgThrThrCysCys 75
Db 1104 CAGACCACTTGTGTGCGCCCGAGCTGCTGCCACCCAGTCTGTGTGTCAGACCACC---TGC 1160
QY 76 ArgProSerCysCysIleSerSerCysCysArgProSerCysCys----- 90
Db 1161 CGCCCCAGCTGTGTGTCTCCAGCTGCTGCCGCCACCTCTGTCAGACCACCTGCCAC 1220
QY 91 -----IleSerSerCysCysLysProSerCysCysArgThrThrCysCysArg 106
Db 1221 CCCAGCTGTGTATGTCAGCTGCTGCCGTCCTCACTCTGTGTGTCAGACCACC---TGC 1277
QY 107 ProSerCysCysIleSerSerCysCysArgProSerCysCysValleSerSerCysCysLys 126
```

```

Db 1278 CCAGCTGTGGTGTGCCAGTGTCTGCCGTCCACTCTGTCTGCAGACCACCTGTCTGCCGT 1337
Qy 127 ProSerCysArgThrThr----- 133
Db 1338 GCAACTTGTCTGCCGCCAGCTGTCTGGATCCTCTTGTGAACTCATATTGGACTATC 1397
Qy 134 -----CysCysArgProSer----- 138
Db 1398 AACCATGAGCCAGTCACCATCCATGATGATGAAGAAGTGTCTGCCGTCCAGCTGTCTGC 1457
Qy 139 -----CysCysLeuSerSerCysCysArgProSerCysCysLeuSerSerCysCys 155
Db 1458 CAGACCACCTGTCTGCAGGACCACTGTCTGCCGCCAGCTGTCTGTGTCTCCAGCTGTCTGC 1517
Qy 156 LysProSerCysCysGlnThrThrCysCysArgProSerCysCysLeuSerSerCysCys 175
Db 1518 AGACCGCAGTGTCTGCCAGTGTCTGTCTGCCAGCCACCTGTCTGCCAGCCAGCTGTCTGC 1577
Qy 176 ArgProGlnCysCysGlnProSerCysCysArgProAlaCysCysLeuSerSerCysCys 195
Db 1578 CAGACCACCTGTCTGCAGGACCACTGTCTGCCGTCTCCAGCTGTCTGTCTCCAGCTGTCTC 1637
Qy 196 HisProSerCysCysValSer-----SerCysArgCysProPheSerCys 210
Db 1638 AGACCCAGTGTCTGCCAGTGTCTGTCTGTCTGCCAGCCACCTGTCTGCCGCCAGCTGTCTGC 1697
Qy 211 ProThrThrCysCysArgThrThrCysCysPheHisProLeuSerSerCysCys 230
Db 1698 CAGACCACCTGTCTGCAGGACCACTGTCTGCCGCCAGCCAGCTGTCTGTCTCCAGCTGTCTGC 1757

RESULT 7
AAK52135
ID AAK52135 standard; cDNA; 1851 BP.
XX
AC AAK52135;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 680.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
XX
PR 27-APR-2000; 2000US-0560875.
XX
PR 20-JUN-2000; 2000US-0598075.
XX
PR 19-JUL-2000; 2000US-0620325.
XX
PR 01-SEP-2000; 2000US-0654936.
XX
PR 15-SEP-2000; 2000US-0663561.
XX
PR 20-OCT-2000; 2000US-0693325.
XX
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR P-PSDB; AAM79002.
XX
Nucleic acids encoding polypeptides with cytokine-like activities,

```

```

PT useful in diagnosis and gene therapy -
XX Claim 1; Page 2374-2376; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM8020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1851 BP; 379 A; 558 C; 574 G; 340 T; 0 other;

Alignment Scores:
Pred. No.: 4,43e-27 Length: 1851
Score: 767.00 Matches: 130
Percent Similarity: 58.80% Conservative: 37
Best Local Similarity: 45.77% Mismatches: 55
Query Match: 51.68% Indels: 62
DB: 22 Gaps: 13

US-09-874-062-3 (1-230) x AAK52135 (1-1851)
Qy 6 CysGlySer-----ValCysSer-----AepGln 13
Db 940 TGGCGCAGCAGCAGCGGTCTTGCAGCAGTGGTCTGGCAGCAGCTGGATGGCAGCAGA 999
Qy 14 SerCysGlyGlnGlyLeuGlyGlnGluSer----- 23
Db 1000 TCTCTGGCAGAGGTCTTAACACACACCTCTCTGAACAGAGAGCCACCTTACCCTCGAC 1059
Qy 24 -----CysCysArgProSerCysCysGlnThrThrCysCys 35
Db 1060 ACCATGACCCACTGTCTTCCCTTGTCTGCTGACCTATCTGTCAGGACCACTGCTGC 1119
Qy 36 ArgThrThrCysCysArgPro-----SerCysCysLeuSerSerCysCysArg 51
Db 1120 AGGACAACTGCTGGAAGCCCACTGTGACCACTGTGACCACTGTGACCACTGTGACCACT 1179
Qy 52 ProSerCysCysLeuSerSerCysCysLys-----ProSerCysCysLeuThr 67
Db 1180 CCTCTCTGTGTGCCAGCTGTCTGCCAGCTTGTCTGCCACCACTTGTCTGCTCAAAAC 1239
Qy 68 ThrCysCysArgThrThrCysCysArgProSerCysCysLeuSerSerCysCysArgPro 87
Db 1240 ACCTGCTGCAGGACCACTGTCTGCCAGCCACT---TGTGTGCCAGCTGTCTGCCAGCT 1296
Qy 88 SerCysCysLeuSerSerCysCysLysProSerCysCysArgThrThrCysCysArgPro 107
Db 1297 TCTGCTGAGCAGCAGCCCTGTGCGCAGCCACCTGCTGTGGGTCCAGCTGTCTGTCGCA 1356
Qy 108 SerCysCysLeuSerSerCysCysArgProSerCysCys-----CysLeuSerSerCys 124
Db 1357 ACCAGCTGTGGTCCAGCTGTCTGAGCCTATTGTGGTCCAGTGTCTGCTGCTGCTGCTGC 1416
Qy 125 CysLysProSerCys-----Cys---ArgThrThrCysCysArgProSerCys 139
Db 1417 TGTCAACCCGACTGTGTATCAAACTATCTGCTTTCAGGACCACTGCTGTCAGCTGTCTGC 1476
Qy 140 CysLeuSerSerCysCysArg-----ProSerCysCysLeuSerSerCys 154
Db 1477 TGGCAGCCCACTGTCTGAGGAAACACCTTCTGCCAGCCACCTGCTGTGGGTCCAGCTGC 1536
Qy 155 CysLysProSerCys-----CysGlnThrThrCysCysArg 166

```

Db 1537 TGCAGCTTGTGCGCCACCAACATGCTGTCAACCAATTGTAGATCCACCTGCTGCCAA 1596
 Qy 167 ProSerCysCysIleSerSerCysCysArgProGlnCysCysGlnProSerCysCysArg 186
 Db 1597 CCATCTGTGTGACCATGCTGCAGCACACCC---TGTGCCACCAACCTGTGTGGG 1653
 Qy 187 ProIlaCysCysIleSerSerCysCysHisProSerCysCysValSerSerCysCysArg 206
 Db 1654 TCCAGCTGCTGTAGCCAAACCTGCAATGAGTCCAGCTATTGTCTGCTTGTGCGCT--- 1710
 Qy 207 ProPheSerCysProThrThrCysCysArgThrThrCysCysPheHisProIleCysCysGly 226
 Db 1711 CCCACCTGCTGCCAGACCACTGTCTACAGGACCACTGTGCTGCCGCCACGCTGTGTGCTGC 1770
 Qy 227 SerSerCysCys 230
 Db 1771 AGTCCTGCTGT 1782

RESULT 8
 AAK53119/c
 ID AAK53119 standard; cDNA; 1851 BP.
 XX
 AC AAK53119;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 2648.
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US04098.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang B, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 WPI; 2001-476283/51.
 DR P-PSDB; AAM79986.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 1; Page 4903; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX
 SQ Sequence 1851 BP; 340 A; 574 C; 558 G; 379 T; 0 other;
 Alignment Scores:
 Pred. No.: 4,436-27 Length: 1851
 Score: 767.00 Matches: 130
 Percent Similarity: 58.80% Conservative: 37
 Best Local Similarity: 45.77% Mismatches: 55
 Query Match: 51.68% Indels: 62
 DB: 22 Gaps: 13

US-09-874-062-3 (1-230) x AAK53119 (1-1851)

Qy 6 CysGlySer-----ValCysSer-----AspGln 13
 Db 912 TGGGGCAGCAGCAAGCGGTCTTGCAGCAGGTGTCTGGCAGCAGCTGGATGGCAGAGA 853
 Qy 14 SerCysGlyGlnGlyLeuGlyGlnGlySer----- 23
 Db 852 TCTTGGCAGAGGTCTTAACCAACACCTCTCTCAACAGGAGCCACCCCTTACCCCTGAC 793
 Qy 24 -----CysCysArgProSerCysCysGlnThrThrCysCys 35
 Db 792 ACCATGACCCACTGCTGTTCCCTTGTGTGTCAGCTACATGTCGACGACCACTGTGTC 733
 Qy 36 ArgThrThrCysCysArgPro-----SerCysCysIleSerSerCysCysArg 51
 Db 732 AGGACAACTGCTGGAAGCCCAACCACTGTGACCACTGTGACGACCAACCCCTGTGTCAG 673
 Qy 52 ProSerCysCysIleSerSerCysCysLys-----ProSerCysCysLeuThr 67
 Db 672 CCCTCTGCTGTGTCAGCTGCTGTCAGCTTGTGTCAGCACTGCTGTGTCAGCACTGCTGT 613
 Qy 68 ThrCysCysArgThrThrCysCysArgProSerCysCysIleSerSerCysCysArgPro 87
 Db 612 ACCTGTGTCAGCACCCTGCTGTCAGCCCACT---TGTGTGGCAGCTGTGTCAGCCT 556
 Qy 88 SerCysCysIleSerSerCysCysLysProSerCysCysArgThrThrCysCysArgPro 107
 Db 555 TCCTGTGTCAGCACACCTGCTGTCAGCCCACTGCTGTGGTCCAGCTGTGTCAGCACTGTC 496
 Qy 108 SerCysCysIleSerSerCysCysArgProSerCysCys-----CysIleSerSerCys 124
 Db 495 ACCAGCTGTGGTCCAGCTGCTGTGACGCTATTGTGGGTCCAGTTGCTGTGTCAGCTTC 436
 Qy 125 CysLysProSerCys-----Cys-----ArgThrThrCysCysArgProSerCys 139
 Db 435 TGTCAACCGACTGCTATCAAACTATCTGCTTCAAGGACCACTGCTGTCAGCCTACCTGTC 376
 Qy 140 CysIleSerSerCysCysArg-----ProSerCysCysIleSerSerCys 154
 Db 375 TGCAGGCCCACTGCTGCGAGGAACACCTCTTGGCAGCCCACTGCTGTGGGTCCAGCTGC 316
 Qy 155 CysLysProSerCys-----CysGlnThrThrCysCysArg 166
 Db 315 TGCAGCCTGCTGTCAGCCCACTGCTGTCAAACTATTGTAGATCCACCTGTCAGCA 256
 Qy 167 ProSerCysCysIleSerSerCysCysArgProGlnCysCysGlnProSerCysCysArg 186
 Db 255 CCATCTGTGTGACCATGCTGCAGCACACCC---TGTGCCAGCAACCTGTGTGGTGGG 199
 Qy 187 ProIlaCysCysIleSerSerCysCysHisProSerCysCysValSerSerCysCysArg 206
 Db 198 TCCAGCTGCTGTAGCCAAACCTGCAATGAGTCCAGCTATTGTGCTGCTGTGCTGCTGCT 142
 Qy 207 ProPheSerCysProThrThrCysCysArgThrThrCysCysPheHisProIleCysCysGly 226
 Db 141 CCCACCTGCTGCCAGACCACTGTCTACAGGACCACTGTTGCCGCCACCACTGTGTGCTGC 82

US-09-874-062-3 (1-230) x ABA09521 (1-410)

```

QY 1 MetValSerSerCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGly 20
Db |||||
48 ATGGTCAGCTCCTGTTGTGGCTCCGTTGTGCTGACCAAGGGCTGGGGCCAAAGACCTCTGT 107
QY 21 GlnGluSerCysArgProSerCysGlnThrThrCysCysArgThrThrCysCys 40
Db |||||
108 CAGGAGACCTGTGCGCCCGCCAGCTGCTGTGAGACCACTGTGAGACCACTGTGCTGC 167
QY 41 ArgProSerCysCysIleSerSerCysArgProSerCysCysIleSerSerCysCys 60
Db |||||
168 CGCCCCAGCTGTTGTATCCAGCTGCTGACGGCCCACTGTGCTGCTGTGCTGC 227
QY 61 -----LysProSerCysCysLeuThrThrCysCysArgThrThrCysCys 75
Db |||||
228 CAACCCACTTGTCCCGCCCGCCAGCTGCTGACCACTGTGACCACTGTGCTGCTAC 287
QY 76 ArgProSerCysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCys 95
Db |||||
288 CGCCCCAGCTGTTGTGTCAGCTGCTGACGGCC-----CAGTGTGCTGC 332
QY 96 LysProSerCysCysArgThrThrCysCysArgProSerCysCysIleSerSerCysCys 115
Db |||||
333 CAGCCTGTGTGTCACCACTGCTGCTGCCCCAGCTGCTGAGACGACCTGCTGC 392
QY 116 ArgProSerCysCysIle 121
Db |||||
393 CACCCTANGTGTGCATC 410

```

Search completed: December 13, 2003, 05:52:29
Job time : 338 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 13, 2003, 05:17:28 ; Search time 3730 Seconds
(without alignments)
2522.578 Million cell updates/sec

Title: US-09-874-062-3
Perfect score: 1484
Sequence: 1 MVSSCGSVCSQSCQGLG.....PTTCRCRTTCFHPICGSSCC 230

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US09874062/runat_12122003_171322_24931/app_query.fasta_1.391
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09874062@cgn_1_3508@runat_12122003_171322_24931 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb_ba:
2: gb_htg:
3: gb_in:
4: gb_om:
5: gb_ov:
6: gb_pat:
7: gb_ph:
8: gb_pl:
9: gb_pr:
10: gb_ro:
11: gb_sta:
12: gb_sy:
13: gb_un:
14: gb_vi:
15: gb_ba:
16: em_fun:
17: em_hun:
18: em_in:
19: em_mu:
20: em_om:
21: em_or:
22: em_ov:
23: em_pat:
24: em_ph:
25: em_pl:
26: em_ro:
27: em_sta:
28: em_un:

29: em_vi:
30: em_htg_hum:
31: em_htg_inv:
32: em_htg_other:
33: em_htg_mus:
34: em_htg_pln:
35: em_htg_rod:
36: em_htg_mam:
37: em_htg_vrt:
38: em_sy:
39: em_htgo_hum:
40: em_htgo_mus:
41: em_htgo_other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1269.5	85.5	347572	2	AC107303	AC107303 Homo sapi
2	1229.5	82.9	161987	9	AC006070	AC006070 Homo sapi
3	1145.5	77.2	221952	2	AC098902	AC098902 Rattus no
4	1145.5	77.2	256796	2	AC099183	AC099183 Rattus no
5	1131.5	76.2	197658	2	AL646094	AL646094 Mus muscu
6	1131.5	76.2	200602	10	AL590997	AL590997 Mouse DNA
7	1110	74.8	197658	2	AL646094	AL646094 Mus muscu
8	1110	74.8	200602	10	AL590997	AL590997 Mouse DNA
9	1006.5	67.8	118444	10	AL590992	AL590992 Mouse DNA
10	996.5	67.1	192027	2	AC067715	AC067715 Mus muscu
11	982	66.2	174032	2	AC025904	AC025904 Homo sapi
12	982	66.2	347572	2	AC107303	AC107303 Homo sapi
13	981.5	66.1	221952	2	AC098902	AC098902 Rattus no
14	977	65.8	1203	9	HSAA06939	AJ406939 Homo sapi
15	977	65.8	186078	9	AC037482	AC037482 Homo sapi
16	970.5	65.4	148801	9	AC100808	AC100808 Homo sapi
17	949	63.9	55568	2	AC015850_3	Continuation (4 of
18	930.5	62.7	174032	2	AC025904	AC025904 Homo sapi
19	926.5	62.4	1080	9	BC004180	BC004180 Homo sapi
20	926.5	62.4	1080	9	BC004212	BC004212 Homo sapi
21	926.5	62.4	1093	9	HSAA06943	AJ406943 Homo sapi
22	926.5	62.4	148801	9	AC100808	AC100808 Homo sapi
23	926.5	62.4	186078	9	AC037482	AC037482 Homo sapi
24	924.5	62.3	1099	9	HSAA06936	AJ406936 Homo sapi
25	908.5	61.2	1220	9	HSAA06944	AJ406944 Homo sapi
26	895.5	60.3	262533	2	AC095873	AC095873 Rattus no
27	882	59.4	262533	2	AC095873	AC095873 Rattus no
28	876.5	59.1	1064	9	HSAA06941	AJ406941 Homo sapi
29	858.5	57.9	1037	9	HSAA06945	AJ406945 Homo sapi
30	838	56.5	142565	10	AL592545	AL592545 Mouse DNA
31	817.5	55.1	110000	2	AC015850_0	AC015850 Homo sapi
32	808.5	54.5	972	10	BC016249	BC016249 Mus muscu
33	805	54.2	118444	10	AL590992	AL590992 Mouse DNA
34	804.5	54.2	942	9	HSAA06937	AJ406937 Homo sapi
35	804	54.2	63280	2	AC105960	AC105960 Mus muscu
36	804	54.2	238286	2	AC122604	AC122604 Rattus no
37	804	54.2	238343	2	AC123463	AC123463 Rattus no
38	804	54.2	245472	2	AC137473	AC137473 Rattus no
39	795.5	53.6	235727	2	AC114148	AC114148 Rattus no
40	775.5	52.3	110000	2	AC015850_0	AC015850 Homo sapi
41	771.5	52.0	271642	2	AC114133	AC114133 Rattus no
42	763.5	51.4	832	11	BV073113	BV073113 S212P6794
43	762	51.3	924	9	HSAA069168	AJ296168 Homo sapi
44	758.5	51.1	161987	9	AC006070	AC006070 Homo sapi
45	734.5	49.5	148047	10	AL731864	AL731864 Mouse DNA

ALIGNMENTS

RESULT 1

AC107303
LOCUS Homo sapiens chromosome 3 clone RP11-595I22, WORKING DRAFT
DEFINITION 347572 bp DNA linear HTG 09-MAY-2002
SEQUENCE, 32 unordered pieces.
AC107303
ACCESSION HTG; HTGS_PHASE1; HTGS_DRAFT.
VERSION AC107303.3 GI:20335965
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 347572)
AUTHORS Murzyn D.M., Adams C., Adio-Oduola B., Ali-Osman P.R., Allen C.,
Albrooks S.L., Anaratunge H.C., Are J.R., Ayele M., Banks T.,
Barbaria J., Benton J., Blmage K., Blankenburg K., Bonnin D.,
Bowick J., Bowie S., Brivea M., Brown E., Brown M., Bryant N.P.,
Buhay C., Burch P., Burkett C., Burrell K.L., Byrd N.C.,
Carroll T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,
Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C.,
Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,
Davila M.L., Davis C., Davy-Carroll L., Dederich D.A.,
Delaney K.R., Delgado O., Denn A.L., Ding Y., Dinh H.H.,
Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.J.,
Earnhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,
Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,
Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,
Gorrell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,
Harris C., Harris K., Hart M., Havlak P., Hawes A., Hernandez J.,
Hernandez O., Hodgson A., Hogue M., Holloway C., Hollins B.,
Honsi F., Howard S., Huber J., Hulyk S., Hume J., Jackson L.E.,
Jacobson B., Jia Y., Johnson R., Jolivet S., Joudah S.,
Karissom E., Kelly S., Khan U., King L., Korvah J., Kovar C.,
Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C., Lewis L.,
Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W., Louisseged H.,
Lozano R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J.,
Maheshwari M., Mapa P., Martin R., Martindale A., Martinez E.,
Massey E., Mawhiney E., McLeod M.P., Meador M., Mei G., Metzker M.,
Miner G., Miner Z., Mitchell T., Mohabbat K., Morgan M., Morris S.,
Moser M., Neal D., Newton J., Newton N., Nguyen A., Nguyen N.,
Nguyen N., Nickerson E., Nwokenkwo S., Ogih M., Okunolu G.,
Oragunye N., Oviado R., Pace A., Payton B., Peery J., Perez L.,
Peters L., Pickens R., Primus E., Pu L.L., Quiles M., Ren Y.,
Rives M., Rojas A., Rojibokan I., Rolfe M., Ruiz S., Savery G.,
Scherer S., Scott G., Shen H., Shoostari N., Sisson I.,
Sodergren E., Sonaite T., Sparks A., Stanley H., Stone H.,
Sutton A., Svatek A., Tabor P., Tamerisa A., Tamerisa K., Tang H.,
Tansey J., Taylor C., Taylor T., Telford B., Thomas N., Thomas S.,
Usmani K., Vaquez L., Vera V., Villalón D., Vinson R., Wang Q.,
Wang S., Ward-Moore S., Warren R., Washington C., Watlington S.,
Williams G., Williamson A., Wleczek R., Wooden S., Worley K.,
Wu C., Wu Y., Wu Y.F., Zhou J., Zorrilla S., Nelson D.,
Weinstock G. and Gibbs R.
Direct Submission
Unpublished
2 (bases 1 to 347572)
Worley K.C.
Direct Submission
TITLE Submitted (18-JAN-2002) Human Genome Sequencing Center, Department
JOURNAL of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 347572)
Worley K.C.
Direct Submission
TITLE Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
JOURNAL of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT
On Apr 28, 2002 this sequence version replaced gi:18449841.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project information

Center project name: HDSS
Center clone name: RP11-595I22
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 31666 bases at least Q40
Consensus quality: 319305 bases at least Q30
Consensus quality: 344140 bases at least Q20
Estimated insert size: 172965; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 2351: contig of 2351 bp in length
* 2352: gap of unknown length
* 2451: contig of 3741 bp in length
* 6192: gap of unknown length
* 6193: contig of 4047 bp in length
* 10333: gap of unknown length
* 10340: contig of 3242 bp in length
* 13681: gap of unknown length
* 13682: contig of 4512 bp in length
* 18293: gap of unknown length
* 18294: contig of 4541 bp in length
* 22933: gap of unknown length
* 22935: contig of 3571 bp in length
* 26605: gap of unknown length
* 26606: contig of 4925 bp in length
* 26706: gap of unknown length
* 31631: contig of 5267 bp in length
* 31731: gap of unknown length
* 36997: contig of 4534 bp in length
* 37097: gap of unknown length
* 41631: contig of 4534 bp in length
* 41731: gap of unknown length
* 41732: contig of 6721 bp in length
* 48453: gap of unknown length
* 48553: contig of 7254 bp in length
* 55806: gap of unknown length
* 55907: contig of 9282 bp in length
* 55908: gap of unknown length
* 65288: contig of 6569 bp in length
* 71857: gap of unknown length
* 71858: contig of 8555 bp in length
* 71958: gap of unknown length
* 80513: contig of 7037 bp in length
* 80613: gap of unknown length
* 87650: contig of 6373 bp in length
* 87750: gap of unknown length
* 94123: contig of 8842 bp in length
* 94223: gap of unknown length
* 103064: contig of 8895 bp in length
* 103164: gap of unknown length
* 103165: contig of 8895 bp in length
* 112059: gap of unknown length
* 112159: gap of unknown length
* 120884: contig of 8725 bp in length
* 120984: gap of unknown length
* 134158: contig of 13174 bp in length
* 134159: gap of unknown length
* 146831: contig of 12573 bp in length
* 146931: gap of unknown length
* 158213: contig of 11282 bp in length
* 158214: gap of unknown length
* 158314: contig of 13888 bp in length
* 172201: gap of unknown length
* 172301: contig of 11927 bp in length
* 184228: gap of unknown length
* 184229: contig of 11170 bp in length
* 195498: contig of 11170 bp in length


```

/repeat_region /map="17"
/repeat_region /clone_lib="RPCI-11 human BAC library"
/repeat_region complement(29..218)
/repeat_region /rpt_family="L2"
/repeat_region complement(260..512)
/repeat_region /rpt_family="L2"
/repeat_region complement(800..1013)
/repeat_region /rpt_family="MER20"
/repeat_region 1140..1416
/repeat_region /rpt_family="AluSx"
/repeat_region 1419..1453
/repeat_region /rpt_family="AluSx"
/repeat_region complement(1476..1515)
/repeat_region /rpt_family="CA)n"
/repeat_region complement(1476..1515)
/repeat_region /rpt_family="MIR"
/repeat_region 1589..1770
/repeat_region /rpt_family="L1ME3"
/repeat_region 1786..1905
/repeat_region /rpt_family="TA)n"
/repeat_region complement(2657..2784)
/repeat_region /rpt_family="MIR"
/repeat_region 3298..3422
/repeat_region /rpt_family="MIR"
/repeat_region complement(6000..6123)
/repeat_region /rpt_family="AluSp/q"
/repeat_region complement(6127..6238)
/repeat_region /rpt_family="LTR41"
/repeat_region 6640..6751
/repeat_region /rpt_family="CA)n"
/repeat_region 7330..7398
/repeat_region /rpt_family="MIR"
/repeat_region 7759..7899
/repeat_region /rpt_family="L1ME2"
/repeat_region complement(7909..8115)
/repeat_region /rpt_family="MERS8A"
/repeat_region complement(8116..8903)
/repeat_region /rpt_family="LTR1"
/repeat_region complement(8904..8928)
/repeat_region /rpt_family="MERS8A"
/repeat_region 8935..9031
/repeat_region /rpt_family="L1ME2"
/repeat_region 9060..9363
/repeat_region /rpt_family="L1ME2"
/repeat_region 10471..10518
/repeat_region /rpt_family="AT rich"
/repeat_region complement(10646..11003)
/repeat_region /rpt_family="MLT1F"
/repeat_region 11004..11209
/repeat_region /rpt_family="L1PA5"
/repeat_region 11218..11296
/repeat_region /rpt_family="TA)n"
/repeat_region 11326..11561
/repeat_region /rpt_family="THE1B"
/repeat_region complement(11779..12083)
/repeat_region /rpt_family="AluSg"
/repeat_region complement(12319..12588)
/repeat_region /rpt_family="MLT1-INTERNAL"
/repeat_region complement(12942..13139)
/repeat_region /rpt_family="L2"
/repeat_region complement(13361..13649)
/repeat_region /rpt_family="AluSx"
/repeat_region complement(14208..14490)
/repeat_region /rpt_family="AluY"
/repeat_region complement(14636..14925)
/repeat_region /rpt_family="AluSg"
/repeat_region 16356..16481
/repeat_region /rpt_family="AluSg"
/repeat_region 16482..16555
/repeat_region /rpt_family="CAAA)n"
/repeat_region 16556..16727
/repeat_region /rpt_family="AluSg"
/repeat_region complement(16870..16897)
/repeat_region /rpt_family="CAAAA)n"

```

```

repeat_region complement(17088..17221)
repeat_region /rpt_family="AluSq/x"
repeat_region complement(17989..18273)
repeat_region /rpt_family="AluSx"
repeat_region 18409..18438
repeat_region /rpt_family="AT rich"
repeat_region complement(18439..18723)
repeat_region /rpt_family="AluSx"
repeat_region 19095..19584
repeat_region /rpt_family="MER31B"
repeat_region 20277..20562
repeat_region /rpt_family="AluSc"
repeat_region 22621..22641
repeat_region /rpt_family="AT rich"
repeat_region 23843..23992
repeat_region /rpt_family="L2"
repeat_region 24184..24291
repeat_region /rpt_family="MIR"
repeat_region 25327..25351
repeat_region /rpt_family="AT rich"
repeat_region complement(25352..25907)
repeat_region /rpt_family="L1PA2"
repeat_region complement(27038..28106)
repeat_region /rpt_family="L1MA4A"
repeat_region complement(28116..28209)
repeat_region /rpt_family="MLT1C"
repeat_region complement(28515..28581)
repeat_region /rpt_family="purine-rich"
repeat_region complement(28582..28674)
repeat_region /rpt_family="MLT1C"
repeat_region 29080..29354
repeat_region /rpt_family="LTR16C"
repeat_region 29355..29510
repeat_region /rpt_family="MIR"
repeat_region 29524..29569
repeat_region /rpt_family="AT rich"
repeat_region complement(29595..29863)
repeat_region /rpt_family="L1MA9"
repeat_region complement(30929..31490)
repeat_region /rpt_family="L1M3e"
repeat_region 32091..32199
repeat_region /rpt_family="MER81"
repeat_region complement(32754..33081)
repeat_region /rpt_family="MLT1B"
repeat_region 33645..33759
repeat_region /rpt_family="LTR16C"
repeat_region 33760..33915
repeat_region /rpt_family="MIR"
repeat_region 33929..33971
repeat_region /rpt_family="AT rich"
repeat_region complement(33997..34259)
repeat_region /rpt_family="L1MA9"
repeat_region complement(35330..35875)
repeat_region /rpt_family="L1M3e"
repeat_region 36529..36637
repeat_region /rpt_family="MER81"
repeat_region complement(37178..37548)
repeat_region /rpt_family="MLT1B"
repeat_region complement(37860..38117)
repeat_region /rpt_family="polypurine"
repeat_region complement(38090..38209)
repeat_region /rpt_family="purine-rich"
repeat_region 38432..38695

```

Alignment Scores:

Pred. No.:	1.28e-36	Length:	161987
Score:	1229.50	Matches:	195
Percent Similarity:	84.78%	Conservative:	0
Best Local Similarity:	84.78%	Mismatches:	0
Query Match:	82.85%	Indels:	35
DB:	9	Gaps:	1

US-09-874-062-3 (1-230) x AC006070 (1-161987)


```

Qy 201 ValSerSerCysArgCysProPheSerCysProThrThrCysArgThrThrCysPhe 220
Db 23617 ATCTCAGTGTGTCGCCGCCATCTGTTGCGAGACCACTGCTGAGGACCACTGCTGCG 23676

Qy 221 HisProIleCysGlySerSerCysCys 230
Db 23677 CCCCAGCATGCTCCAGTGTCTTGTCTGC 23706

RESULT 5
AL646094/c 197658 bp DNA linear HTG 27-FEB-2002
LOCUS Mus musculus chromosome 11 clone RP23-142E7, *** SEQUENCING IN
DEFINITION PROGRESS ***
ACCESSION AL646094
VERSION AL646094.6 GI:19031714
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Ellington, A.
AUTHORS Direct Submission
TITLE Submitted (26-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 28, 2002 this sequence version replaced gi:18135257.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM142E7
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 99% of reads
Chemistry: Dye-primer Big Dye; 0% of reads
Consensus quality: 192882 bases at least Q40
Consensus quality: 192997 bases at least Q30
Consensus quality: 193039 bases at least Q20
Insert size: 197658; sum-of-contigs
Insert size: 203092; 2.3% error; agarose-fp
Quality coverage: 13.20x in Q20 bases; sum-of-contigs Quality
coverage: 13.17x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..197658
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-142E7"
/clone_lib="RPCI-23"
misc_feature
1..197658
/note="assembly fragment:02832"

BASE COUNT 59207 a 42073 c 41216 g 55162 t
ORIGIN

Alignment Scores:
Pred. No.: 6.2e-33 Length: 197658
Percent: 1131.50 Matches: 172
Conservative: 27
Best Local Similarity: 58.50% Mismatches: 26
Query Match: 76.25% Indels: 69
DB: 2 Gaps: 5

US-09-874-062-3 (1-230) x AL646094 (1-197658)

```

```

Qy 1 MetValSerSerCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGly 20
Db 5336 ATGTTAGTTCCTGTTGTTGCTCTGCTGCTGAGAGGGCTGTGCGCAGGC----- 5283

Qy 21 GlnIleSerCysCysArgProSerCysCysGlnThrThrCysCysArgThrThrCysCys 40
Db 5282 -----TGCTGCCAGCCAGCTGCTGCCAGACCACTGCTGTAGGACCACTGCTGCTGC 5232

Qy 41 ArgProSerCysCysIleSerSer----- 48
Db 5231 CGCCCCAGCTGCTGTGTCCAGCTGCTGCAGACCCAGTGTGCCAGTGTGTGTGCTGC 5172

Qy 49 -----CysCysArg 51
Db 5171 CAGCCCACTGCTGCTGCCCCAGCTGCTGCTATTCTAGCTGTTGTAGACCTTGTGCGCG 5112

Qy 52 ProSerCysCysIleSerSerCysCysLeuProSerCysCysLeuThrThrCysCysArg 71
Db 5111 CCCACCTGCTGTGTGCCAGCTGCTGCAGGCCCTTCTTGTGCTGCAATTCTAGCTGCTGAGG 5052

Qy 72 ThrThrCysCysArgProSerCysCysIleSerSerCysCysArgProSerCysCysIle 91
Db 5051 CTTTCTCTGCTGCCGCCAGCTGCTGTGTGTGCCAGCTGCTGCAGGCCCTCAGTGTGCGCAG 4992

Qy 92 SerSerCysCysLysPro-----SerCysCysArg 101
Db 4991 TCTCTGTGCTGCCAGCCCACTGCTGCCGCCAGCTGTTGCTGCAATTCTAGCTGCTGCAGG 4932

Qy 102 ThrThrCysCysArgProSerCysCysIleSerSerCysCysArgProSerCysCysIle 121
Db 4931 CTTTCTCTGCTGCCGCCAGCTGCTGTGTGTGCCAGCTGCTGCAGGCCCTTCTGCTGCATC 4872

Qy 122 SerSerCysCysLysPro----- 127
Db 4871 TCCAGCTGCTGTGCCGCCCTGCTGTGTGTTCTAGTGTTCGTGATCCAGTGTGCTGC 4812

Qy 128 -----SerCysCysArgThrThrCysCysArgProSerCysCysIle 141
Db 4811 CCAAGCTGTTGCATTTCTAGCTGCTGCAGGGCTTCTTGTGCTGCCGCCCTAGTGTGCTGTA 4752

Qy 142 SerSerCysCysArgProSerCysCysIleSerSerCysCysLysProSerCysCysGln 161
Db 4751 TCCAGCTGCTGCCGCCAGTTGCTGTGTGCCAGCTGCTGCAGACCCAGTGTGCTGCCAG 4692

Qy 162 ThrThrCysCysArgProSerCysCysIleSerSerCysCysArgPro----- 177
Db 4691 TCTCTGTGCTGCCAGCCCACTTGTGCCGCCAGCTGCTGCCGCCAGTGTGTCATT 4632

Qy 178 ---GlnCysCysGlnProSerCysCysArgProAlaCysCysIleSerSerCysCysHis 196
Db 4631 TCTAGCTGCTGCAGACCTTCTTGTGCCGCCAGCTGCTGTGTGTCAGACCACTTGTGTCAGA 4572

Qy 197 ProSerCysCysValSerSerCysArgCysProPheSerCysProThrThrCysCysArg 216
Db 4571 CCCCAGTGTGTATCTCCAGTGTGTCGCCGCCAGCTTGTGTCGAGACCACTTGTGTCAGA 4512

Qy 217 ThrThrCysPheHisProIleCysCysGlySerSerCysCys 230
Db 4511 ACCACCTGTTGTGCCGCCAGCATGCTCTAGTGTGTTCTTGTGCTGC 4470

RESULT 6
AL590997/c 200602 bp DNA linear ROD 29-NOV-2001
LOCUS Mouse DNA sequence from clone RP23-132J20 on chromosome 11,
DEFINITION complete sequence.
ACCESSION AL590997
VERSION AL590997.13 GI:17221205
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```



```

Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BM142E7
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 9% of reads
Chemistry: Dye-primer Big Dye; 0% of reads
Consensus quality: 192882 bases at least Q40
Consensus quality: 192997 bases at least Q30
Consensus quality: 193039 bases at least Q20
Insert size: 197658; sum-of-contigs
Insert size: 203092; 2.3% error; agarose-fp
Quality coverage: 13.20x in Q20 bases; sum-of-contigs Quality
coverage: 13.17x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
    source
        1..197658
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /chromosome="11"
            /clone="RP23-142E7"
            /clone_lib="RPCI-23"
        misc_feature
            1..197658
                /note="assembly fragment:02832"
BASE COUNT 59207 a 42073 c 41216 g 55162 t
ORIGIN
Alignment Scores:
Pred. No.: 3.87e-32 Length: 197658
Score: 1110.00 Matches: 167
Percent Similarity: 77.91% Conservative: 27
Best Local Similarity: 67.07% Mismatches: 31
Query Match: 74.80% Indels: 24
DB: 2 Gaps: 5
US-09-874-062-3 (1-230) x AL646094 (1-197658)
Qy 1 MetValSerSerCysGlySerValCysSerAspGlnSerCysGlyClnIleuGly 20
Db 17739 ATGGTCAGCTCTGTTGTGGCTCTGCTCTCTGAGGAGGGCTGTGGCAAGGC----- 17792
Qy 21 GlnGluSerCysCysArgProSerCysCysGlnThrCysCysArgThrThrCysCys 40
Db 17793 -----TGCTGCCAGCCAGCTGCTGCCAGACTACTGCTGTAGGACCACTGCTGC 17843
Qy 41 ArgProSerCysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCys 60
Db 17844 CCCCCAGCTGCTGTGTGTCCAGCTGCTGCAGCCAGCTGCTGTGTCCAGCTGCTGTC 17903
Qy 61 LysProSerCysCysLeuThr-----ThrCysCysArgThrThrCysCys 75
Db 17904 AGACCCCACTGTTCAGCTCTGTGTGCTGCCAGCCACCTGCTGCCGCCAGCTGCTGT 17963
Qy 76 ArgProSerCysCysIleSerSer-----CysCysArgProSerCysCysIle 91
Db 17964 CGCCCCAGCTGCTGCAATTTCTAGCTGCTGTAGACCTTGTGCGCCGCCAGCTGCTGTGTG 18023
Qy 92 SerSerCysCysAlaProSerCysCysArgThrThrCysCysArgProSerCysCysIle 111
Db 18024 TCCAGCTGTGAGGCCCCAGCTGCTGCCAGTCTGTGTGCTGCCAGCCCACTGCTGCCGC 18083
Qy 112 SerSerCysCysArgProSerCysCysIleSerSerCysCysLysPro----- 127
Db 18084 CCCAGCTGTGTGCGCCGCCAGCTGTTGCATTTCCAGCTGTGCGCCGCCCTGCTGTGTACT 18143

```

```

Qy 128 ---SerCysCysArgThrThrCysCysArgProSerCysCysIleSer----- 142
Db 18144 TCCAGCTGCTGTGTGATCTAGCTGCTGCCGCCACCTGTTGTCATTTCCAGCTGCTGCCGC 18203
Qy 143 ---SerCysCysArgProSerCysCysIleSerSerCysCysCysProSerCysCysGln 161
Db 18204 CCCAGCTGCTGCCAGGCCAGCTGCTGTGTGTCCAGCTGCTGCCAGACCCAGTGTGCATC 18263
Qy 162 ThrThrCysCysArgProSerCysCysIleSerSerCysCysTyrArgProGlnCysCysGln 181
Db 18264 TCCAGCTGCTGCCGCCCTGCTGCTGTGTAGTCTTCTAGCTGCTGTGATCTAGCTGCTGCCGC 18323
Qy 182 ProSerCysCysArgProAlaCysCysIleSerSerCysCysHisProSerCysCysVal 201
Db 18324 CCCAGCTGCTGCCAGGCCAGCTGCTGTGTGTCCAGCTGCTGCCAGACCCAGTGTGCATC 18383
Qy 202 SerSerCysCysArgProSerCysCysProThrThrCysCysArgThrThrCysCysPheHis 221
Db 18384 TCTAGCTGCTGCCGCCACCTGTTGCCAGACCCACCTGCTGCCAGGACCACTTGTGTGCCGC 18443
Qy 222 ProIleCysCysGlySerSerCysCys 230
Db 18444 CCATCATGCTCTAGTGTCTTCTGCTGC 18470

```

RESULT 8

AL590997

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

US-09-874-062-3 (1-230) x AL590992 (1-118444)

```
QY      1 MetValSerSerCysCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGly 20
DB      6214 ATGGTCAGCTCTGTGTGGCTCTGTCTGCTGAGAGAGGGCTGTGGCCAAAGGC----- 6161

QY      21 GlnGluSerCysCysArgProSerCysCysGlnThrThrCysCysArgThrThrCysCys 40
DB      6160 -----TGTTCAGCCAGCTGCTGCCAGACCACTGCTGTAGACCACTGCTGTC 6110

QY      41 ArgProSerCysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCys 60
DB      6109 CGCCCCAGCTGCTGTGTCCAGCTGCTGCAGACCCAGTGTGTCGAGCTGTGTGTGCTGTC 6050

QY      61 LysProSerCysCysLeuThrThrCysCysArgThrThrCysCysArgProSerCysCys 80
DB      6049 CAGCCCC-----ACTGTCTGTGCCCCAGCTGCTGTC 6020

QY      81 IleSerSerCysCysArgProSerCysCysIleSerSerCysCysIleSerProSerCysCys 100
DB      6019 ATCTCCAGCTGTGTGCGCCAGCTGCTGTAGACCACTGCTGTGAGGCCACCTGCTGT 5960

QY      101 ArgThrThrCysCysArgProSerCysCysIleSerSerCysCysArgProSerCysCys 120
DB      5959 AGGCTTCTGCTGCCAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5900

QY      121 IleSerSerCysCysIleSerSerCysCysArgThrThrCysCysArgProSerCysCys 140
DB      5899 GTGTCCAGCTGCTGCCAGGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5849

QY      141 IleSerSerCysCysArgProSerCysCysIleSerSerCysCysIleSerProSerCysCys 160
DB      5848 -----ACCTGCTGCTGCCAGCTGCTGCTATCTAGTTGTCGCTCCTAGCTGCTGT 5795

QY      161 GlnThrThrCysCysArgProSerCysCysIleSerSerCysCysTyrArgProGlnCysCys 180
DB      5794 GTGTCCAGCTGCTGCCAGGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5747

QY      181 GlnProSerCysCysArgProAlaCysCysIleSerSerCysCysIleSerProSerCysCys 200
DB      5746 GGGTCTAGCTGCTGCCAGGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5687

QY      201 ValSerSerCysCysArgProSerCysCysProThrThrCysCysArgThrThrCysCysPhe 220
DB      5686 GTGTCCAGCTGCT-----TGCAGGCCACCTGCTGCTGCCAGACCTGCTGCTGT 5642

QY      221 HisProIleCysCysGlySerSerCysCys 230
DB      5641 CGCCCAACTGCTCTAGTTGTTCTGTTGCTGTC 5612
```

RESULT 10

```
AC067715 192027 bp DNA linear HTG 07-JAN-2001
LOCUS     Mus musculus chromosome 11 clone RP23-326H14, WORKING DRAFT
DEFINITION
```

```
AC067715 192027 bp DNA linear HTG 07-JAN-2001
VERSION   HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS  Mus musculus (house mouse)
```

```
SOURCE   Mus musculus
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
(bases 1 to 192027)
```

```
Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
Gorell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,J., Liu,W., Louissegh,H., Lozada,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
```

Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogih,M., Parish,B.,
Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
Williamson,A., Wrenford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstein,G.,
Worley,K. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 192027)
Worley,K.C.
Direct Submission
Submitted (27-APR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 4, 2001 this sequence version replaced gi:1192108.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: MACE

Center clone name: RP23-326H14

----- Summary Statistics

Sequencing vector: M13; 108821

Chemistry: Dye-terminator Big Dye: 57% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 189676 bases at least Q40

Consensus quality: 191575 bases at least Q30

Consensus quality: 192401 bases at least Q20

Estimated insert size: 189593; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 7.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 6 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence.

* as soon as it is available and the accession number will

* be preserved.

* 1 71311: contig of 71311 bp in length

* 71312 71411: gap of unknown length

* 71412 116268: contig of 44857 bp in length

* 116269 116368: gap of unknown length

* 116369 145312: contig of 28944 bp in length

* 145313 145412: gap of unknown length

* 145413 173610: contig of 28198 bp in length

* 173611 173710: gap of unknown length

* 173711 190868: contig of 17158 bp in length

* 190869 190968: gap of unknown length

* 190969 192027: contig of 1059 bp in length.

* Location/Qualifiers

1..192027

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="11"

/clone="RP23-326H14"

BASE COUNT 55026 a 41185 c 39554 g 55749 t 513 others

ORIGIN

Alignment Scores:

Pred. No.: 68-28 Length: 192027

Score: 996.50 Matches: 155

Percent Similarity: 74.35% Conservativeness: 16

Best Local Similarity: 67.33% Mismatches: 24

Query Match: 67.15% Indels: 35

DB: 2 Gaps: 5

* 74185 74284: gap of 100 bp
 * 74285 86712: contig of 12428 bp in length
 * 86713 86812: gap of 100 bp
 * 86813 103527: contig of 18715 bp in length
 * 103528 103627: gap of 100 bp
 * 103628 120465: contig of 16838 bp in length
 * 120466 120565: gap of 100 bp
 * 120566 145510: contig of 24945 bp in length
 * 145511 145610: gap of 100 bp
 * 145611 174032: contig of 28422 bp in length.

FEATURES

Location/Qualifiers
 1..174032
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="RP11-585G19"
 /clone_lib="RPC1-11 Human Male BAC"

misc_feature
 1..2588
 /note="assembly_fragment"
 clone end:77
 vector side:right

misc_feature
 2689..6465
 /note="assembly_fragment"

misc_feature
 6566..8846
 /note="assembly_fragment"

misc_feature
 8947..13104
 /note="assembly_fragment"

misc_feature
 13205..19430
 /note="assembly_fragment"

misc_feature
 19531..25564
 /note="assembly_fragment"

misc_feature
 25665..31377
 /note="assembly_fragment"

misc_feature
 31478..39523
 /note="assembly_fragment"

misc_feature
 39624..47042
 /note="assembly_fragment"

misc_feature
 47143..55411
 /note="assembly_fragment"

misc_feature
 55512..66000
 /note="assembly_fragment"

misc_feature
 66101..74184
 /note="assembly_fragment"

misc_feature
 74285..86712
 /note="assembly_fragment"

misc_feature
 86813..103527
 /note="assembly_fragment"

misc_feature
 103628..120465
 /note="assembly_fragment"

misc_feature
 120566..145510
 /note="assembly_fragment"

misc_feature
 145611..174032
 /note="assembly_fragment"
 clone end:SP6
 vector side:right

BASE COUNT 50722 a 34581 c 34553 g 52576 t 1600 others
 ORIGIN

Alignment Scores:
 Pred. No.: 1..92e-27 Length: 174032
 Score: 982.00 Matches: 157
 Percent Similarity: 74.68% Conservative: 17
 Best Local Similarity: 67.38% Mismatches: 35
 Query Match: 66.17% Indels: 24
 DB: 2 Gaps: 6

US-09-874-062-3 (1-230) x AC025904 (1-174032)

Qy 1 MetValSerSerCysCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGly 20
 |||||
 Db 143089 ATGGTCAGCTCCTGTTGTGGCTCCGTTGCTCTGACACAGGCGTGCAGGCTCTCTGT 143030
 |||||
 Qy 21 GlnGluSerCysCysArgProSerCysGlnThrThrCysCysArgThrThrCysCys 40

Db 143029 CAGGAGACCTGCTGCGCGCCAGCTGCTGTGCAGACCACTGTTGCAGGACCACTGCTAC 142970
 |||||:::|||||
 Qy 41 ArgProSerCysCysValleSerSerCysCysArgProSerCysCysValleSerSerCysCys 60
 |||||
 Db 142969 CGCCCAAGCTGTTGTGTGCTCAGCTGCTGCAGGCGCCAGTGTGCTGCTGCTGCTGTC 142910
 |||||
 Qy 61 LysProSerCysCysLeuThrThrCysCysArgThrThrCysCysArgProSerCysCys 80
 :::::|||||
 Db 142909 CAACCCACCTGCTGCTGCCCGCAGCTGCTGTGAGACGACCTGCTGCCACCCCTAGTGTGCTGC 142850
 |||||
 Qy 81 IleSerSerCysCysArgProSerCysCysValleSerSerCysCysValleSerProSerCysCys 100
 |||||
 Db 142849 ATCTCCAGCTGCTGCGCGCCAGCTGCTGTATGTCAGCTGCTGCAGGCCCACTGCTGC 142790
 |||||
 Qy 101 ArgThr-----ThrCysCysArgProSerCysCysValleSerSerCysCys 115
 :::::|||||
 Db 142789 CAGTCTGTGCTGCCAGCCCACTGCTGCGCGCCAGCTGCTGCATCTCCAGCTGCTGT 142730
 |||||
 Qy 116 ArgProSerCysCysValleSerSerCysCysLysProSerCysCysArgThr----- 132
 |||||
 Db 142729 CGCCCAAGCTGCTGTGCTCAGGTGCTGCAGGCGCCAGTGTGCCAGTCTGTGTGCTGC 142670
 |||||
 Qy 133 -----ThrCysCysArgProSerCysCysValleSerSerCysCysArgProSerCysCys 150
 |||||
 Db 142669 CAGCCAACTGCTGCCGCTCCAGCTGCTGCATCTCCAGCTGCTGCCGCCCTCTGTGCTGT 142610
 |||||
 Qy 151 IleSerSerCysCysLysProSerCysCysGlnThrThrCysCys---ArgProSerCysCys 169
 |||||
 Db 142609 GAATCCAGCTGCTGCCGCCCACTGCTGTGCGGCCCTGCTGCTGCTGCCAGTGTGCTGC 142550
 |||||
 Qy 170 -----CysIleSerSerCysThrArgProGlnCysCysGlnProSerCysCys 185
 |||||
 Db 142549 GGCCGAGTCTCTGCGCACCACTGCTATCGCCCACTGCTGTCATCTCCACCTGTGCC 142490
 |||||
 Qy 186 ArgProAlaCysCysValleSerSerCysCys-----HisPro 197
 |||||
 Db 142489 CGCCCTTGTGCTGTGCTCTCTTGTGCTGAGCCCACTGCCCTGCTGCTGCTGCTGCTGCT 142430
 |||||
 Qy 198 SerCysCysValSerSerCysArgProPheSerCys 210
 |||||
 Db 142429 TCACCACTGGCCCACTGATGAGA---CCCTTCTACTGT 142394
 |||||

RESULT 12

AC107303/c

LOCUS

DEFINITION

SEQUENCE, 32 unordered pieces.

AC107303

AC107303.3 GI:20335965

VERSION

HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 347572)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

Albrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,

Barbaria,J., Benton,J., Binagge,K., Blankenburg,K., Bonnin,D.,

Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.F.,

Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,

Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,

Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinu,H.H.,

Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,

Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,

Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,

Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,

Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kuresni, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, W., Louleaged, H.,
 Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sison, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 347572)
 Worley, K.C.
 Direct Submission
 Submitted (18-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 347572)
 Worley, K.C.
 Direct Submission
 Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Apr 28, 2002 this sequence version replaced gi:18449841.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HDSS
 Center clone name: RP11-595122
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 33166 bases at least Q40
 Consensus quality: 339305 bases at least Q30
 Consensus quality: 344140 bases at least Q20
 Estimated insert size: 172965; sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see <http://www.hgsc.bcm.tmc.edu/docs/genbankdraft.data.html>).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 32 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 2351: contig of 2351 bp in length
 * 2352 2451: gap of unknown length
 * 2452 6192: contig of 3741 bp in length
 * 6193 6292: gap of unknown length
 * 6293 10339: contig of 4047 bp in length
 * 10340 10439: gap of unknown length
 * 10440 13681: contig of 3242 bp in length

13682 13781: gap of unknown length
 13782 18293: contig of 4512 bp in length
 18294 18394: gap of unknown length
 18395 22334: contig of 4541 bp in length
 22335 22935: gap of unknown length
 22936 26005: contig of 3571 bp in length
 26006 26705: gap of unknown length
 26706 31630: contig of 4925 bp in length
 31631 31731: gap of unknown length
 31732 36997: contig of 5267 bp in length
 36998 37097: gap of unknown length
 37098 41631: contig of 4534 bp in length
 41632 41731: gap of unknown length
 41732 48452: contig of 6721 bp in length
 48453 48552: gap of unknown length
 48553 55806: contig of 7254 bp in length
 55807 55906: gap of unknown length
 55908 65188: contig of 9282 bp in length
 65189 65288: gap of unknown length
 65289 71957: contig of 6569 bp in length
 71958 80512: contig of 8555 bp in length
 80513 80612: gap of unknown length
 80613 87649: contig of 7037 bp in length
 87650 87749: gap of unknown length
 87750 94122: contig of 6373 bp in length
 94123 94222: gap of unknown length
 94223 103664: contig of 8842 bp in length
 103665 103664: gap of unknown length
 103665 112059: contig of 8895 bp in length
 112060 112059: gap of unknown length
 112061 120884: contig of 8725 bp in length
 120885 120884: gap of unknown length
 120885 134158: contig of 13174 bp in length
 134159 134259: gap of unknown length
 134260 146831: contig of 12573 bp in length
 146832 146831: gap of unknown length
 146833 158213: contig of 11282 bp in length
 158214 158313: gap of unknown length
 158314 172201: contig of 13888 bp in length
 172202 172301: gap of unknown length
 172302 184328: contig of 11927 bp in length
 184329 184328: gap of unknown length
 184330 195498: contig of 11170 bp in length
 195499 195598: gap of unknown length
 195600 210747: contig of 15148 bp in length
 210748 210847: gap of unknown length
 210848 226329: contig of 15482 bp in length
 226330 226429: gap of unknown length
 226430 246181: contig of 19753 bp in length
 246182 246281: gap of unknown length
 246282 265510: contig of 19229 bp in length
 265511 265610: gap of unknown length
 265611 293686: contig of 28076 bp in length
 293687 293786: gap of unknown length
 293787 347572: contig of 53786 bp in length.
 ----- Location/Qualifiers
 1. 347572
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-595122"
 BASE COUNT 107326 a 64859 c 63333 g 108900 t 3154 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,14e-27 Length: 347572
 Score: 982.00 Matches: 157
 Percent Similarity: 74.68% Conservative: 17
 Best Local Similarity: 67.38% Mismatches: 35
 Query Match: 66.17% Indels: 24
 DB: 2 Gaps: 6

US-09-874-062-3 (1-230) x AC107303 (1-347572)

Qy 1 MetValSerSerCysCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGly 20
 Db 311429 ATGGTCAGCTCTGTTGTGGCTCCGTCGTCTGACACAGGGCTGGGCAAGTCTCTGT 311370
 Qy 21 GlnGluSerCysCysArgProSerCysCysGlnThrThrCysCysArgThrThrCysCys 40
 Db 311369 CAGGAGACCTGCTGCGCCCGCCAGCTGCTGTGACACACCTGTTGACAGACCACTGCTAC 311310
 Qy 41 ArgProSerCysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCys 60
 Db 311309 CGCCCCAGCTGTTGTGTCACAGCTGCTGACAGCCCGCCAGTGTGCCAGTGTGTGCTGTC 311250
 Qy 61 LysProSerCysCysLeuThrThrCysCysArgThrThrCysCysArgProSerCysCys 80
 Db 311249 CAACCCACCTGCTGTCGCCCGCCAGCTGCTGTGACAGACCTGCTGTCACCCCTAGTGTGTC 311190
 Qy 81 IleSerSerCysCysArgProSerCysCysIleSerSerCysCysIleSerSerCysCys 100
 Db 311189 ATCTCCAGCTGCTGCGCCCGCCAGCTGCTGTATGTCAGCTGCTGTCAGCCCGCCAGTGTGTC 311130
 Qy 101 ArgThr-----ThrCysCysArgProSerCysCysIleSerSerCysCys 115
 Db 311129 CAGTCTGTGCTGTCGACGCCACCTGCTGCGCCCGCCAGTGTGTCATCTCCAGCTGCTGT 311070
 Qy 116 ArgProSerCysCysIleSerSerCysCysLysProSerCysCysArgThr----- 132
 Db 311069 CGCCCCAGCTGCTGTGTCAGCTGCTGTCAGCCCGCCAGTGTGTCAGTCTGTCGCTGTC 311010
 Qy 133 -----ThrCysCysArgProSerCysCysIleSerSerCysCysArgProSerCysCys 150
 Db 311009 CAGCCCAACTGCTGCGCGCCCGCCAGTGTGTCATCTCCAGCTGCTGCGCCCGCCCTGTCGT 310950
 Qy 151 IleSerSerCysCysLysProSerCysCysGlnThrThrCysCys---ArgProSerCys 169
 Db 310949 GAATCAGCTGCTGCGCCCGCCAGTGTGTCGCGCCCGCTGCTGCTGCTGCTGCTGCTGCT 310890
 Qy 170 -----CysIleSerSerCysCysArgProGlnCysCysGlnProSerCysCys 185
 Db 310889 GGCGGAGTCTCTGCGCACACCACTGCTATCGCCCAACCTGTCATCTCCACCTGTCCTCC 310830
 Qy 186 ArgProAlaCysCysIleSerSerCysCys-----HisPro 197
 Db 310829 CGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 310770
 Qy 198 SerCysCysValSerSerCysCysArgCysProPheSerCys 210
 Db 310769 TCACCACTGCCACAGATGTAGA---CCCTTCTACTGT 310734

RESULT 13

AC098902/c
 LOCUS 221952 bp DNA linear HTG 10-MAY-2003
 DEFINITION Rattus norvegicus clone CH230-166D21, WORKING DRAFT SEQUENCE.

AC098902

AC098902.7 GI:30521241

HTG; HTGS PHASE2; HTGS DRAFT; HTGS FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

AUTHORS

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, P.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
 Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flegg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Grægeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idiebird, D., Jackson, A.,
 Jackson, B., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowitz, C., Kraft, C. L., Lebow, H., Levay, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorenschew, L., Loulseg, H., Lozado, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Manthey, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwokeneme, O., Okwum, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Poinexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smaj, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
 Weinstein, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 221952)

Worley, K. C.

Direct Submission

Submitted (05-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 221952)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Alignment Scores:
 Pred. DOI: 8.7e-29 Length: 1203
 Score: 977.00 Matches: 156
 Percent Similarity: 74.25% Conservative: 17
 Best Local Similarity: 66.95% Mismatches: 36
 Query Match: 65.84% Indels: 24
 DB: 9 Gaps: 6

US-09-874-062-3 (1-230) x HSA406939 (1-1203)

Qy	1	MetValSerSerCysCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGly	20
Db	47	ATGGTCAGCTCTGTTGTGGCTCCGTGCTCTGACCCAGGGCTGCGGGCAAGCTCTCTGT	106
Qy	21	GlnGluSerCysCysArgProSerCysCysGlnThrThrCysCysArgThrThrCysCys	40
Db	107	CAGGAGACCTGCTGCGCGCCAGCTGCTGTCAGACCACTGTTGTCAGGACCACTCTCTAC	166
Qy	41	ArgProSerCysCysIleSerSerCysCysArgProSerCysCysIleSerSerCysCys	60
Db	167	CGCCCCAGCTGTTGTGTCAGCTGCTGCGAGGCCAGTGTGCCAGTCTGTGTGCTGTCG	226
Qy	61	LysProSerCysCysLeuThrThrCysCysArgThrThrCysCysArgProSerCysCys	80
Db	227	CAACCCACCTGCTGTCGCCCGCCAGCTGCTGAGACGACCTGCTGCCACCCCTAGGTGCTGC	286
Qy	81	IleSerSerCysCysArgProSerCysCysIleSerSerCysCysLysLysProSerCysCys	100
Db	287	ATCTCCAGCTGCTGCGCGCCAGCTGCTGTATGTCTGACGCTGCTCAAGCCCCAGTGTGC	346
Qy	101	ArgThr-----ThrCysCysArgProSerCysCysIleSerSerCysCys	115
Db	347	CAGTCTGTGTGTCGCCCGCCAGCTGCTGCCACCCAGCTGCTGCATCTCCAGTGTCTGT	406
Qy	116	ArgProSerCysCysIleSerSerCysCysLysProSerCysCysArgThr-----	132
Db	407	CGCCCCAGCTGCTGTGTGTCAGGCTGCTGCAGGCCAGCTGCTGCCAGTCTGTGTGCTGC	466
Qy	133	-----ThrCysCysArgProSerCysCysIleSerSerCysCysArgProSerCysCys	150
Db	467	CAGCCAACTGCTGCGCTGCCAGCTGCTGCATCTCCAGCTGCTGCGGCCCTCTTGTCTGT	526
Qy	151	IleSerSerCysCysLysProSerCysCysGlnThrThrCysCys---ArgProSerCys	169
Db	527	GAATTCAGCTGCTGCGCCCATGCTGCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTG	586
Qy	170	-----CysIleSerSerCysCysArgProGlnCysCysGlnProSerCysCys	185
Db	587	GGCCGAGTCTCTGCCACACCACTGCTATGCCCAACCTGTGTGTCATCTCCACCTGCTCC	646
Qy	186	ArgProLysCysCysIleSerSerCysCys-----HsaPro 197	
Db	647	CGCCCCCTGTGCTGTGCTCTCTCTGCTGAGCCCACTGCGCTGCTGCTGCTGCTGCTGCT	706
Qy	198	SerCysCysValSerSerCysCysArgCysProPheSerCys 210	
Db	707	TCACCACTGGCCACAGATGTAGA---CCCTTCTACTGT 742	

RESULT 15
 AC037482/c
 LOCUS Homo sapiens chromosome 17, clone RP11-619M15, complete sequence.
 DEFINITION
 AC037482
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 186078)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE
 Homo sapiens chromosome 17, clone RP11-619M15

Unpublished
2 (bases 1 to 186078)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczyk, J., Levine, R., Lie, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., MCPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., P. Pierre, N., Peterson, K., P. Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (09-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 186078)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczyk, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (16-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 186078)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczyk, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

REFERENCE
AUTHORS

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 TITLE
 JOURNAL
 Submitted (23-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Feb 23, 2002 this sequence version replaced gi:18693527.
 COMMENT
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WTBH
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L9334
 Center clone name: 619_M_15

Only the first 186.1 kilobases of this clone are being submitted.
 The remainder overlaps accession number AC004231 [WICGR project L297].

FEATURES

```

source
1..186078
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="17"
   /map="17"
   /clones="RP11-619M15"
   /clone_lib="RPC1-11 Human Male BAC"
   complement(1..313)
   /rpt_family="L1"
   complement(314..642)
   /rpt_family="L1MC5"
   complement(1000..1127)
   /rpt_family="L1MC4a"
   1134..11465
   /rpt_family="AluSg1"
   complement(1504..1585)
   /rpt_family="L1MA9"
   complement(1707..2219)
   /rpt_family="L1MC4a"
   complement(2278..2566)
   /rpt_family="AluSx"
   complement(2682..2944)
   /rpt_family="L1PA8"
   2953..2983
   /rpt_family="AT-rich"
   3070..3231
   /rpt_family="MIR"
   4320..4877
   /rpt_family="L1M2b"
   complement(5609..6042)
   /rpt_family="L1PA16"
   6499..6580
   /rpt_family="TTCCCN"
   /note="single clone coverage"
   6581..6704
   /rpt_family="L1R16A"
   complement(7231..7419)
   /rpt_family="L1PA5"
   7597..7776
   /rpt_family="L1R16C"
   complement(8190..8513)
   /rpt_family="L2"
   8562..8751
   /rpt_family="MIR"
   9055..9346
   /rpt_family="AluSc"
   10091..10547
   /rpt_family="L1M2b"

repeat_region
complement(11689..11981)
/rpt_family="L1R43"

repeat_region
complement(12328..12677)
/rpt_family="MER4A"

repeat_region
complement(12680..12776)
/rpt_family="MER4A"

repeat_region
complement(12780..13499)
/rpt_family="MER4A-int"

repeat_region
complement(13500..13622)
/rpt_family="L1R1B"

repeat_region
complement(13700..14460)
/rpt_family="L1R1B"

repeat_region
complement(14461..16352)
/rpt_family="MER4-int"

repeat_region
complement(16352..17004)
/rpt_family="MER4-int"

repeat_region
complement(17004..17056)
/rpt_family="MER4-int"

repeat_region
17057..17295
/rpt_family="L1R8"

repeat_region
complement(17296..17983)
/rpt_family="MER4A-int"

repeat_region
18006..18044
/rpt_family="(CA)n"

repeat_region
complement(18046..18244)
/rpt_family="MER4A-int"

repeat_region
18245..18404
/rpt_family="(TTTC)n"

repeat_region
complement(18418..18977)
/rpt_family="L1PA7"

repeat_region
complement(19023..19236)
/rpt_family="MER4A-int"

repeat_region
complement(19234..19392)
/rpt_family="MER4A-int"

repeat_region
complement(19476..19715)
/rpt_family="MER4A-int"

repeat_region
complement(19864..20487)
/rpt_family="MER4A-int"

repeat_region
complement(20491..20862)
/rpt_family="MER4A"

repeat_region
complement(20863..20961)
/rpt_family="MER4A"

Alignment Scores:
Pred. No.: 3,09e-27 Length: 186078
Score: 977.00 Matches: 156
Percent Similarity: 74.25% Conservative: 17
Best Local Similarity: 66.95% Mismatches: 36
Query Match: 65.84% Indels: 24
DB: 9 Gaps: 6

US-09-874-062-3 (1-230) x AC037482 (1-186078)

Qy 1 MetValSerCysCysGlySerValCysSerAspGlnSerCysGlyGlnGlyLeuGly 20
Db 44421 ATGGTCAGCTCCTGTTGTGGCTCCGTGCTGCTGACAGGGCTGGCGCCAGGCTCTGT 44362

Qy 21 GlnGluSerCysCysArgProSerCysCysGlnThrThrCysCysArgThrThrCysCys 40
Db 44361 CAGGAGACCTGCTGCCGCCAGCTGCTGTGCAGACCACTGTTGTCAGGACCACTGCTAC 44302

Qy 41 ArgProSerCysCysAlaSerSerCysCysArgProSerCysCysAlaSerSerCysCys 60
Db 44301 CGCCCCAGCTGTTGTGTGTCAGCTGCTGCAGGCCCGCCAGTGTGTCAGTGTGTCGTC 44242

Qy 61 LysProSerCysCysLeuThrThrCysCysArgThrThrCysCysArgProSerCysCys 80
Db 44241 CAACCCACCTGCTGTGCCCGCCAGCTGCTGTGAGACGACCTGTGCCACCTAGTGCTGC 44182

Qy 81 IleSerSerCysCysArgProSerCysCysAlaSerSerCysCysAlaSerProSerCysCys 100
Db 44181 ATCTCCAGCTGCTGCCGCCAGCTGCTGTATGTCCAGCTGCTGCAAGCCCGAGTGTGCTGC 44122

```

